UNIVERSITY OF BAMBERG
CUMULATIVE DISSERTATION THESIS

The Creation, Formalization, and Transfer of Expert Knowledge with Visual Analytics in Industrial Manufacturing Processes of Electrical Vehicles

Author:
Joscha EIRICH

Supervisor:
Prof. Dr. Ute Schmid *External Reviewer:*
Prof. Dr. Tobias Schreck
Graz University of Technology
Further Committee Members:
Prof. Dr. Kai Fischbach

Date of defense: January 16, 2023

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Natural Sciences (Dr. rer. nat)

Faculty for Information Systems and Applied Computer Sciences
University of Bamberg

November 21, 2022

Bamberg, 2023

URN: urn:nbn:de:bvb:473-irb-583392
DOI: https://doi.org/10.20378/irb-58339

Declaration of Authorship

I, Joscha EIRICH, declare that this thesis titled, “The Creation, Formalization, and Transfer of Expert Knowledge with Visual Analytics in Industrial Manufacturing Processes of Electrical Vehicles” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"Imagination is more important than knowledge, because knowledge is limited."

Albert Einstein

To my Mom and Dad

Acknowledgements

This thesis is a result of the continuous encouragement and support I have received from a wide network of brilliant researchers and practitioners. First of all, I thank BMW for the extraordinary opportunity to do this thesis within an organizational and industrial context. This was an enriching and at the same time challenging experience, whereby I was able to address real-world problems with state-of-the-art research methods. Work that emerged during this thesis was very much appreciated and awarded by different research communities as well as by my industrial partners. Of course, this would not have been possible without the support of some great world-leading scientists, who supported me during this thesis.

First of all, I thank Prof. Dr. Ute Schmid, my doctoral advisor at the University of Bamberg, for accepting me as an external doctoral candidate and for the strong support that she provided for my research projects during this thesis. Second, I thank Prof. Dr. Tobias Schreck from the Graz Institute of Technology. He played one of the key roles in the research work carried out during this thesis, and heavily supported my projects with his great body of knowledge about information visualization and his network with other motivated and gifted researchers. With the help of Tobias, we managed to significantly enhance the body of information visualization by demonstrating how applied research can contribute to solving real-world problems. I am honored to have him as my second reviewer and very much hope that our great collaboration will continue after finishing this thesis. I also thank Prof. Dr. Kai Fischbach from the University of Bamberg, for his willingness to be part of the doctoral committee. Another Thanks goes to Dr. Diana Fischer Preßler for her strong emotional support during the first challenging year of my thesis and her support in performing theory-driven research.

Apart from that, a special thanks go to my dear colleague Prof. Dr. Jürgen Bernard from the University of Zürich, for his strong support, the great and honest feedback for many of my research projects, and for the great collaboration we had during the last two years. I also thank Prof. Dr. Michael Sedlmair from the University of Stuttgart for his support and teaching of applied research methodologies as well as his recommendations and feedback on my research works. Furthermore, I would like to thank my colleagues from Pro2Future located in Austria, especially Dr. Belgin Mutlu and Georgios Koutroulis for their fruitful collaboration and support during this thesis.

BMW also played an instrumental role during this initial phase of my research career and throughout this thesis. In this regard, I would like to thank my internal supervisor Dr. Jose Bittencourt for the continuous support in presenting my scientific findings inside BMW, for believing in my work, and for his guidance as a fellow researcher during the last years. I also thank Dr. Thomas Herzinger for always giving me all the freedom I needed to perform my research while balancing the demands for the practical applicability of my work. Another thanks goes to Andre Luckow for his great scientific guidance, believing in my potential, and for promoting my work inside non-manufacturing related departments of BMW. Furthermore, I want to thank Kai Demtröder for being my mentor at BMW and introducing me to other heads of departments to broaden my field of vision inside my organization. Another thanks goes to Dr. Stefan Werrlich for reviewing my research work and for his guidance when I started this doctoral program.

Probably the biggest thanks for this whole thesis, however, goes to Dr. Dominik Jäckle. He taught me all the necessary skills on how to write scientific publications, supported me in solving technical problems, and introduced me to the research field

of Information Visualization. Especially when I started this thesis, I was very doubtful if I would be able to successfully finish this doctoral program. If it was not for Dominik, I probably would have quit this thesis in my first year, where I did not receive the support I needed from other researchers at my University. I cannot thank him enough, for supporting me when my self-confidence took a blow.

I owe what I am in my life to my beloved parents - my mother Ulrike Eirich and my father Andreas Eirich as well as my brother Christopher Eirich and my sister Marena Eirich. Therefore, a special thanks go to my beloved family for believing in me all the time, sticking to me in all the hard times, and being there for me when I needed them most.

This doctoral thesis was a journey of lows and highs, wins and losses. But I believe that every struggle along that road made me a stronger and bolder person, gave me insights about life, and awakened the desire to continuously explore new things. This journey was far more than worth taking, and I am thankful for being fortunate enough to have the opportunity of taking this road towards science. That being said, my final thanks and deep love goes to my partner Anneke Christiansen, who had to endure me when times were rough and always supported and accompanied me in good and bad times. If it would not be for her, I would not be at this point of my life where I am now.

Abstract

The 21st century will be heavily impacted by the capability to create value from recorded data of all kinds. In this regard, industrial manufacturing intuitions increasingly rely on new data-driven technologies, such as the Internet of Things or Machine Learning. In terms of data collection, manufacturing processes increasingly include sophisticated sensor equipment, which results in interconnected networks of manufacturing all assembling parts and producing data. However, manufacturing institutions currently face two challenges. First, large amounts of parts and hence data are produced during fully automated manufacturing processes. Second, due to the overwhelming amount of recorded data, it is particularly challenging to efficiently analyze manufacturing data. Hence, it is important to efficiently store and share gained knowledge from performed data analyses.

Information visualization and Visual Analytics are two prolific branches of data analysis exploiting sophisticated visualization techniques to support the execution of analytical tasks and to store gained knowledge. This thesis looks at how data visualization approaches can help industrial manufacturing organizations to create value from large amounts of manufacturing data, as well as how to efficiently store and share knowledge from manufacturing data analyses. The goal is to understand how Visual Analytics can improve manufacturing processes in the context of knowledge management. Five Visual Analytics systems were designed, developed, and evaluated to tackle different domain problems that emerged from manufacturing setups. Findings from each system were used to carry out additional studies to enhance established theories of knowledge management. As a result, five success stories are provided of how Visual Analytics can significantly improve manufacturing processes and how knowledge can be efficiently created, formalized, and shared in an organization with Visual Analytics.

Zusammenfassung

Der Erfolg vieler Unternehmen im 21. Jahrhundert wird stark davon geprägt sein in wieweit sie in der Lage sein werden aufgezeichnete Daten jeglicher Art wertschöpfend zu nutzen. Besonders industrielle Hersteller setzen dabei zunehmend auf neue datengtriebene Technologie wie beispielsweise das Internet der Dinge oder maschinelles Lernen. Hierbei beinhalten Herstellungsprozesse zunehmend ausgeklügelte Sensorik, welche es ermöglicht Produktionsanlagen miteinander zu verbinden und somit einen Gesamtüberblick über hergestellte Bauteileteile und aufgezeichnete Daten zu ermöglichen. Industrielle Hersteller stehen derzeit jedoch vor zwei Herausforderungen. Erstens werden in vollautomatisierten Fertigungsprozessen große Mengen an Teilen und damit Daten produziert.

Zweitens ist es aufgrund der überwältigenden Menge an aufgezeichneten Daten eine besondere Herausforderung, Fertigungsdaten effizient zu analysieren. In diesem Kontext ist es besonders wichtig, gewonnenes Wissen aus durchgeführten Datenanalysen effizient zu speichern und zu teilen. Hierbei stellen die Diziplin der Informationsvisualisierung und der Visual Analytics zwei Zweige der Datenanalyse dar, welche ausgefeilte Visualisierungstechniken nutzen, um die Ausführung analytischer Aufgaben zu unterstützen und gewonnenes Wissen zu speichern. Im Zuge dieser Promotion wird untersucht, wie Datenvisualisierungsansätze industrielle Fertigungsunternehmen dabei unterstützen können, Mehrwert aus großen Mengen von aufgezeichneten Daten zu schaffen und wie Wissen aus Fertigungsdatenanalysen effizient gespeichert und geteilt werden kann. Das Ziel dabei ist zu verstehen, wie Visual Analytics Fertigungsprozesse im Kontext des Wissensmanagements verbessern kann. Als resultat wurden fünf Visual Analytics-Systeme designed, implementiert und evaluiert, um verschiedene Domänenprobleme, die sich aus Fertigungsumgebungen ergaben zu addressieren. Anhand der Erkenntnisse aus jedem System wurden im Rahmen zusätzlicher Studien etablierte Theorien des Wissensmanagements analyisert und verbessert. Das Ergebnis dieser Arbeit beinhaltet fünf Erfolgsgeschichten wie Visual Analytics Fertigungsprozesse erheblich verbessern kann und wie Wissen mit Visual Analytics in einer Organisation effizient erstellt, formalisiert und geteilt werden kann.

List of Publications

Journal Publications

- **J. Eirich**, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. IRVINE: A Design Study on Analyzing Correlation Patterns of Electrical Engines. *IEEE Transactions on Visualization and Computer Graphics*, pp.11-21, 07 2021.
- **J. Eirich**, G. Koutroulis, B. Mutlu, D. Jäckle, R. Kern, T. Schreck, and J. Bernard. ManEx: The Visual Analysis of Measurements for the Assessment of Errors in Electrical Engines. *IEEE Computer Graphics and Applications*, pp. 1–13, 03 2022.
- **J. Eirich**, M. Münch, D. Jäckle, M. Sedlmair, J. Bonart, and T. Schreck. RfX: A Design Study for the Interactive Exploration of a Random Forest to Enhance Testing Procedures for Electrical Engines. *Computer Graphics Forum*, pp. 1–14, 03 2022.

Conference Publications

- J. Bonart, **J. Eirich**, P. Hümmer, and W.-G. Drossel. Identification of Anomalies in Highly-Integrated Electric Drives by Secondary Excitation Mechanisms. In *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, 08 2021.
- **J. Eirich**. Visual Analytics for IoT Data from Large-Scale Manufacturing processes. In *American Conference on Information Systems*, 08 2022.
- **J. Eirich** and F.-P. Diana. The Life Cycle of Data Labels in Organizational Learning: A Case Study of the Automotive Industry. In *European Conference on Information Systems*, 04 2022.
- **J. Eirich**, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. VIMA: Modeling and Visualization of High Dimensional Machine Sensor Data Leveraging Multiple Sources of Domain Knowledge. In *Visualization in Data Science at IEEE VIS*, 09 2020.
- **J. Eirich**, D. Jäckle, S. Werrlich, and T. Schreck. Visual Analytics in Organizational Knowledge Creation: A Case Study. In *European Conference on Information Systems*, 04 2021.

Book Chapters

Apart from the conference and journal publications, these two book chapters are about to be publicized in 2022.

- **J. Eirich**, A. Luckow, and Demtroeder. K. Visual Analytics zur Optimierung von Pruefprozessen in der Serienfertigung elektrischer Motoren. In *Wie funktionier Data Science?*, pp. 1–7. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

- A. Luckow, **J. Eirich**, and Demtroeder. K. Kuenstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie. In *Wie funktionier Data Science?*, pp. 1–11. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

To be Published

Furthermore, the following publication was submitted to the journal IEEE Transactions on Visualization and Computer Graphics.

- **J. Eirich**, D. Jäckle, M. Sedlmair, C. Wehner, U. Schmid, J. Bernard, T. Schreck, and J. Bernard. ManKnowVis: How to Support Different User Groups in Contextualizing and Leveraging Knowledge Repositories.

Contents

List of Figures

Part I

Part A - Research Outline

Chapter 1

Introduction

Industrial manufacturing institutions across the globe currently face one of the greatest changes in the last 100 years, namely the automation and digitization of all manufacturing processes. In this regard, industrial manufacturing intuitions increasingly rely on new data-driven technologies, such as the Internet of Things (IoT) [134] or Machine Learning [96] to name a few. In terms of data collection, manufacturing processes increasingly include sophisticated sensor equipment, which results in interconnected IoT networks and hence a new generation of smart factories. The "smartness" level in a factory, however, depends to great extent on how organizations can leverage data-driven approaches to create value for the resulting big data sets [63] and how organizations can support human experts with data analysis tasks [32]. However, industrial manufacturers face two main challenges in this transformation.

1) Industrial manufacturing processes produce large numbers of parts resulting in a vast amount of recorded sensor data throughout the whole manufacturing process. The holistic analysis of this data, however, is often overwhelming for often view domain experts with the necessary knowledge to interpret the data correctly [30]. For example, we found that engineering experts were required to analyze more than forty interdependent signals from hundreds of produced electrical engines produced at a single testing station manually and per hour [28].

2) Due to the vast amount of data, only a few produced parts can be analyzed in detail. Hence, the gained knowledge from such in-depth analyses is particularly important to improve the overall quality of manufacturing processes, regarding part testing, manufacturing output, or the identification of errors. Yet, it is currently a tedious process to efficiently store and share gained knowledge from part analyses. For example, one of our studies revealed that for an assembly line of battery modules engineering experts frequently analyze data from six different databases where no system is present, which allows an efficient sharing of gained knowledge from analyzing individual data sources [31].

One way to tackle these problems is to use data visualization approaches [23, 115], which have intrinsic motivation and a long-standing tradition of applied and problem-driven research [72]. The typical goal of such projects is to enable domain experts to understand and analyze their own domain data through carefully designed visualization interfaces [105]. In this context, the discipline of Visual Analytics [115] can help to integrate human experts into the visual exploration of manufacturing data by supporting the analytical reasoning capabilities of experts with model-supported and custom-designed visualization interfaces [57]. Furthermore, Visual Analytics technologies are capable of receiving and storing expert feedback in the form of new data inputs, for example, data labeling or annotation tasks. Hence, Visual Analytics provides an excellent opportunity to tackle the problems of creating value from large amounts of manufacturing data as well as the efficient storing and sharing of knowledge from manufacturing data analyses.

Recent applications of Visual Analytics to manufacturing problems show how this technology can result in substantial manufacturing improvements. For example, Xu et al. [133] show how the visual exploration of an assembly line can help to detect inefficiencies. Fujiwara et al. [39] provide a success story where their Visual Analytics systems help experts in the identification of patterns in machine repair logs to decrease maintenance costs. Sedlmair et al. [104] show how Visual Analytics can help automotive engineers in debugging masses of traces each consisting of millions of recorded messages from in-car communication networks. Finally, Suschnigg et al. [114] show how domain experts can improve the quality of part testing by providing a Visual Analytics system that allows for detecting anomalies in test bench data.

All of the above-mentioned studies are impressive examples, of how Visual Analytics positively impacts the quality of manufacturing processes. However, none of the mentioned studies explicitly reviewed Visual Analytics in the context of how to efficiently formalize and share created knowledge from performed data analyses. This work aims to further investigate this research gap. As a result, we aim to investigate how Visual Analytics can improve manufacturing processes in the context of knowledge management. In so doing, we are one the one side interested in how new knowledge can be created with Visual Analytics in industrial manufacturing settings while we also want to explore how created knowledge from data analysis tasks can be efficiently formalized and shared with Visual Analytics. As a result, we hope to better understand Visual Analytics in the context of industrial manufacturing and how it can be leveraged to efficiently retain knowledge in organizations.

In this thesis, we designed, developed, and evaluated five Visual Analytics systems to tackle different domain problems that emerged from manufacturing setups. Using the findings from each system, we carried out additional studies to further enhance established theories of knowledge management. As a result, we provide five success stories, of how Visual Analytics can significantly improve manufacturing processes and how knowledge can be efficiently created, formalized, and shared in an organization with Visual Analytics.

1.1 Research Questions

Our overall goal in this research is to find answers to the following question:

"How can Visual Analytics support the creation, externalization, and transfer of knowledge in manufacturing processes?"

This question can be divided into three sub-questions. The first research question investigates novel ways to support domain experts with the analysis of large amounts of IoT data from manufacturing processes. In this regard, we want to find out what visualization techniques are adequate when dealing with manufacturing processes with the goal to enable the creation of new knowledge through data analyses. The second research question is about identifying novel ways to capture and formalize created knowledge from previously performed data analyses. This formalized knowledge can be leveraged in multiple ways, which leads us to the next research question. The third research question is about how formalized knowledge can support machines to further support human experts with their analyses. The fourth research question deals with the novel collaboration approaches to sharing formalized expert knowledge among distinct organizational stakeholder groups. In the following, we present a detailed description of each research question:

*"**RQ1:** How can Visual Analytics support domain experts in the analysis of large-scale manufacturing IoT data to create new knowledge?"*

This research question addresses the investigation of algorithms and visualization techniques that support domain experts with the analysis of large-scale IoT data from manufacturing processes. To answer this question, we developed a theoretical framework, which outlines mechanisms of knowledge creation and conversion afforded by visual analytics [32].

Furthermore, we explored possibilities to automatically process manufacturing data to provide users with the most relevant data instances that are adequate for an in-depth analysis. Here, we employed various statistical (e.g., correlation patterns of acoustic signatures [13]) or machine learning (e.g., self-organizing-maps [61, 60, 98]) approaches to recommend users the most relevant data instances for analyses.

To investigate how visualization techniques can support domain experts with their analyses, we deployed and evaluated the two visual analytics systems VIMA (Virtual interactive manufacturing assistant) [30] and IRVINE (Interactive Clustering and Labeling) [28], that process and visualize manufacturing data from electrical engines.

This investigation led to the identification of promising mechanisms to create knowledge with Visual Analytics. For example, we surfaced that the interaction with Visual Analytics systems triggered socialization events between different domain experts, where all involved parties collaboratively generated knowledge. Furthermore, we found out that the computation of anomalies with residual values over large samples of produced parts proved to be an excellent yet comprehensible method to easily and automatically detect interesting parts for an in-depth analysis. Last, we discovered novel ways of visualizing anomalies in manufacturing processes with glyph representations or small multiples as abstractions from acoustic signatures.

*"**RQ2:** How can Visual Analytics support externalizing tacit knowledge."*

This research question focuses on evaluating how generated knowledge from analyses of manufacturing data can be captured and formalized. To answer this question, firstly, we performed background research on mechanisms of organizational knowledge externalization afforded with Visual Analytics [32]. Moreover, we considered previous Visual Analytics applications from the manufacturing sector, to capture and formalize knowledge, for example via state diagrams [104].

Secondly, we developed the Visual Analytics system RfX (Random Forest Explorer) [35]. This Visual Analytics system was designed to support domain experts with the analysis of representative decision trees from a random forest making this black-box machine learning model more understandable.

The investigation from our background research and the applications of our Visual Analytics systems (Vima, IRVINE, RfX) led to important findings on how to capture and formalize gained knowledge from expert analyses. For example, we found out that knowledge can be externalized via labels, rules, or annotations from data analyses, such as rules from representative decision trees of a random forest [35]. Another example is the externalization of expert knowledge via labeling of engines [30], which contain anomalies in their acoustic signatures, and annotating the cause of errors directly in their raw sensor measurements [28]. Last, we contributed to novel ways to interactively analyze a random forest through multiple entry points into analyses, such as icicle plots or two-dimensional representations of decision trees in a random forest.

*"**RQ3:** How can formalized knowledge from Visual Analytics be used to improve a Visual Analytics Systems to further support expert analyses."*

This research question aims to identify ways how Visual Analytics systems can leverage formalized domain knowledge from expert analyses. The goal is to introduce a human-in-the-loop cycle, where domain experts closely work with Visual Analytics systems and both the human expert can perform analyses more efficiently and the system improves over time. In this regard, we particularly performed extensive background research on interactive machine learning [4, 12], interactive labeling [82], and interactive clustering [98] to foster human-machine collaboration.

Second, we deployed and evaluated the Visual Analytics system ManEx [33] that leverages labeled data to identify root causes of erroneous engines in the entire manufacturing process.

The findings that resulted from this research question led to important insights into how humans and machines can collaborate and mutually benefit from each other. For example, in IRVINE we included features to use labeled data to retrain the system's underlying prediction model to provide more detailed suggestions of different error causes for engines. Another example was the demonstration of how causal discovery algorithms from ManEx supported experts in identifying relevant stations across the manufacturing process for an in-depth analysis. Each analysis resulted here in new labels, which again improved the system's underlying causal discovery algorithm. From a visualization point of view, we also contributed novel ways how to integrate causal mapping as a new layer on manufacturing processes and discovered novel ways on comparing different groups of parts with glyph representations.

*"**RQ4:** How can created knowledge efficiently be transferred with Visual Analytics among distinct stakeholder groups involved in manufacturing processes?"*

With this research question, we want to identify novel ways to leverage Visual Analytics to transfer formalized knowledge among distinct organizational stakeholder groups. During the thesis, we noticed a discrepancy between two relevant stakeholder groups, those that provide knowledge, and those that consume it. Knowledge providers, such as engineers, have domain knowledge about specific parts of the manufacturing process. However, they do not have the necessary skills to perform sophisticated data-driven analyses, such as the training of machine learning models. Such analyses are usually performed by highly skilled data experts, such as data scientists. However, this group does not have first-hand domain knowledge about the manufacturing process. Therefore, they rely on consuming knowledge from providers. Hence, based on the previous findings, we wanted to analyze how already formalized knowledge can be efficiently transferred from providers to consumers.

To do so, we first performed a qualitative study to investigate the relation of formalized knowledge in the form of labels to organizational learning [6, 56]. This resulted in a theoretical model of how labels affect organizational learning [29]. Furthermore, we did background research on knowledge-assisted visualization [71, 18, 81], which specifically supports knowledge transfer between stakeholders with different knowledge levels. Based on that, secondly, we developed and evaluated the knowledge-assisted Visual Analytics system ManKnowVis [31].

From this research, we extracted important findings and contributions to better understand how Visual Analytics can be leveraged to foster knowledge transfer. For example, we enhanced traditional theoretical models of organizational learning with labels. Another example is the representation of manufacturing knowledge with an ontology-assisted knowledge graph, which facilitates knowledge sharing via describing and connecting different entities in the knowledge graph. Regarding advances in visualization design, we developed novel visual abstractions to represent knowledge graphs as a semantic layer for manufacturing processes. For example, we provide individual views for different entities in the knowledge graph, such as a representation of manufacturing stations on top of shop floor layouts as freely adaptable rectangles.

1.2 Research Approach

We initiated our research work with literature analyses on multiple topics, such as visual interactive labeling [12], interactive machine learning [4], knowledge-assisted visualization [81], organizational knowledge creation [88], organizational learning [6], knowledge creation in visual analytics [100], or anomaly detection in manufacturing data [114] to name a view. In this regard, we performed two case studies according to the rationale of Yin [135] to enhance the following two theoretical models:

First, we combined the knowledge generation model of Visual Analytics from Sacha et al. [100] with the model of organizational knowledge creation from Nonaka and Takeuchi [88] to derive an integrated theoretical framework to describe organizational knowledge creation and transfer with Visual Analytics.

Second, we extended the model of organizational learning from Argote and Miron Spektor [6] with the concept of labels, to describe how labels-an abstraction of externalized expert knowledge-influence learning processes in organizations.

The theoretical models provided us with the necessary foundation to conceptualize the creation, externalization, and transfer of knowledge through Visual Analytics. In addition to that, we developed five Visual Analytics systems (Vima, RfX, IRVINE, ManEx, ManKnowVis) to demonstrate how Visual Analytics can support knowledge creation, externalization, and transfer in real-world manufacturing scenarios. Each system was developed in the form of a design study according to Sedlmair et al. [105]. Here, we also relied on well-established models that support the design and development of Visual Analytics systems, such as the design triangle of Miksch and Aigner [80], the nested model from Munzer [84], or the visualization mantra of Schneiderman [106].

1.3 Collaboration Statement

This section provides an overview of colleagues and collaborators, who contributed to the work described in this thesis and without whom this thesis would not have been possible.

- Prof. Ute Schmid from the University of Bamberg is the supervisor of this thesis and contributed to the conception of the publications [28, 31] in the paper writing. Furthermore, she contributed to the conception of the research proposal for the public-founded research project KiProQua-A research collaboration between the University of Bamberg and BMW founded by the Bavarian Ministry of Economics.

- Prof. Tobias Schreck from CGV - TU Graz, is the co-reviewer and examiner of this thesis and contributed to the publications [30, 32, 28, 33, 31, 35] in the conception phase and helped with paper writing. He also supported us with important knowledge about the design and development of Visual Analytics systems in the manufacturing sector and provided us with access to Pro2Future-a research institute with a large experience in the application of Visual Analytics in the manufacturing industry.

- Prof. Michael Sedlmair from the University of Stuttgart. He collaborated in the publications [35, 28, 31] with valuable knowledge about human-computer interaction, visual analytics applications in the automotive industry. As one of the pioneers in establishing the methodology of design studies, he also supported us with methodical details on how to perform design studies.

- Prof. Jürgen Bernard from the University of Zurich supported the publications [28, 33, 31]. As one of the developers of the concept of Visual Interactive Labeling [12], he provided us with valuable insight on designing Visual Analytics systems to externalize expert knowledge with labels and the interactive training of machine learning models with labels.

- Prof. Kai Fischbach from the University of Bamberg is the examiner of this thesis and contributed to the publications [30, 28].

- Dr. Andre Luckow from the Ludwig Maximilian University of Munich and head of innovation ad emerging technologies at BMW. He contributed to the publications [73, 34], provided us with access to relevant experts inside BMW, and gave us a platform inside BMW to promote our research.

- Dr. Dominik Jäckle from the BMW Group collaborated in the publications [30, 32, 28, 35, 33, 31] and provided valuable knowledge regarding applications of Visual Analytics and the design and development of all in this thesis presented Visual Analytics Systems.

- Dr. Stefan Werrlich from the BMW Group contributed to the publication [32] and with reviewing the publication [30]. He also provided valuable knowledge in performing case studies.

- Dr. Thomas Herzinger from the BMW Group was the operational supervisor of this thesis. He provided us with access to relevant experts inside BMW and continuously supports the industrialization of the five Visual Analytics systems (Vima, RfX, IRVINE, ManEx, ManKnowVis) inside BMW.

- Dr. Jose Bittencourt from the BMW Group was the mentor for this thesis. He contributed with access to relevant experts inside BMW and continuously supports the industrialization of the five Visual Analytics systems (Vima, RfX, IRVINE, ManEx, ManKnowVis) inside BMW.
- Dr. Belgin Mutlu from Pro2Future supported us with the publication [33] with knowledge about causal discovery and applications of Visual Analytics in industrial settings.
- Dr. Diana Fischer-Preßler from the University of Bamberg contributed to the publication [29]. She supported us with the conceptualization of the paper, coding of expert interviews, and writing of the paper.
- Dr. Roman Kern from TU Graz contributed to the publication [33] with knowledge about the development of causal discovery algorithms and reviewing the paper.
- Georgios Koutroulis from Pro2Future supported us with the publication [33] with the development of its underlying causal discovery algorithm and paper writing.
- Dr. Stefan Scheele from the Fraunhofer IIS Research Group Explainable Artificial Intelligence contributed to the conceptualization of the public founded research project KiProQua-A research collaboration between the University of Bamberg and BMW founded by the Bavarian Ministry of Economics.
- Christoph Wehner from the University of Bamberg supported us in the publication [31] with the development of the system's back-end. Particularly, he developed the system's underlying knowledge graph.
- Jakob Bonart from the BMW Group supported the publications [30, 35, 28] with knowledge about acoustic signatures of electrical engines in serial manufacturing processes. He is also the first author of the paper [13], which forms the physical foundations and conceptualization for the Visual Analytics system IRVINE.
- Markus Münch from Robert Bosch GmbH contributed to the publication [35] with knowledge about the developing syntactic and semantic similarity measurements to compare decision trees of a random forest to identify the most representative decision trees in the random forest.

1.4 Structure of Dissertation

This thesis is composed of two parts: Research Outline (Part A) and Publications (Part B). Part A comprises five chapters.

First, we provide an introduction of this thesis, the research questions (See Section 1.1), a general description of the research approach (See Section 1.2), a list of all collaborators of this thesis 1.3, and a description of the thesis structure 1.4.

The second chapter provides the necessary theoretical foundations for this thesis. Here, we provide an introduction to the area of Visual Analytics (See Section 2.1) and Knowledge Management (See Section 2.2).

In the third chapter, we present our research methodology. First, we describe the design study methodology from Sedlmair et al. [105] (See Section 3.1) and then proceed with additional research methods (See Section 3.2), for example, the design triangle from Miksch and Aigner [80].

The fourth chapter describes the results from this thesis in an Overview of all developed Visual Analytics Systems (See Section 4.1) and developed theoretical models (See Section 4.1).

Finally, we conclude this thesis with a summary and discussion. First, we address how our results address the research questions (See Section 5.1). This is followed by a description of the practical and scientific contributions of this thesis (See Section 5.2), the thesis limitations (See Section 5.3), future work and the conclusion (See Section 5.4).

Part B contains the main publications that lay the core concepts, results, and practical as well as scientific contributions discussed in this thesis.

Chapter 2

Theoretical Background

The focus of this thesis is the study of how Visual Analytics can support the creation, externalization, and transfer of knowledge in industrial manufacturing settings. This thesis considers important conceptual foundations for Visual Analytics, as well as knowledge management. We start with an introduction to Visual Analytics in general, followed by a discussion of existing models for knowledge creation in Visual Analytics. Next, we give a brief overview of the field of Knowledge-Assisted Visualization and conclude with application examples of Visual Analytics systems in the Manufacturing Sector. This is followed by an introduction to knowledge management in general and well-established knowledge management models, approaches, and theories. After that, we summarize the two core knowledge management concepts used in this thesis, which are the SECI model (Socialization, externalization, combination, and internalization) and the process of organizational learning.

2.1 Visual Analytics

The generation and creation of more digital information, for instance, via introducing IoT sensors in manufacturing processes [101], resulted in the need for more efficient data analysis solutions. In this regard, various problems, such as the amount and speed of data generated through high-resolution sensing devices provide no challenges to data analysis. Other problems are related to the filtering, aggregation, and visualization of this data.

In this regard, visualization helps to tell stories by curating data into a form easier to understand for users, who previously were not used to working with data during their daily routines, such as mechanical engineers [28]. A good visualization tells a story, removing the noise from data and highlighting useful information. As part of information visualization, Visual Analytics provides various opportunities to transform data into valuable information and knowledge.

Visual Analytics can be described as "the science of analytical reasoning facilitated by interactive visual interfaces" [115]. By using methods from knowledge discovery in databases, statistics, and artificial intelligence, Visual Analytics systems support human data analysis tasks. During interactions between individuals and the system, the system suggests relevant insights and analytical interfaces, while humans review and revise the systems' output before drawing their conclusions [57]. Figure 2.1 depicts that the scope of Visual Analytics is very broad across multiple research communities, such as knowledge representation, knowledge discovery, or cognitive sciences.

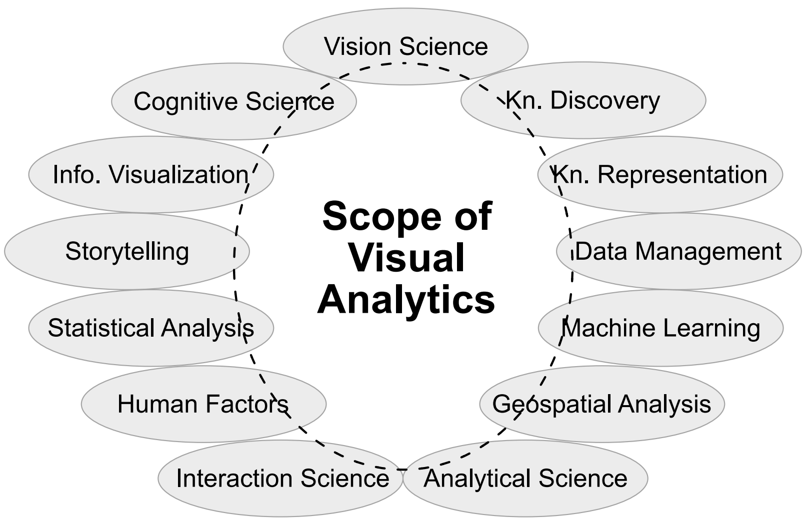


FIGURE 2.1: Visual Analytics scope, according to Keim et al. [57]

Visual Analytics is also closely related to the discipline of Business Intelligence. Similar to VA, Business Intelligence uses predefined methods (e.g., data aggregation or filtering) to gather, analyze, transform, and abstract data into new information via visualization interfaces to inform business decisions [107]. Nonetheless, Visual Analytics Business Intelligence differs in the following aspects: Visual Analytics systems are tailor-made visual interfaces between human experts and machine learning models to support highly explorative analyses. Examples are the analysis of speech data with self-organizing maps [98] or root cause analyses of manufacturing errors with causality graphs [33]. Most Visual Analytics systems are designed to store data in the form of formalized user feedback, which can range from simple text inputs to more abstract forms such as graphs, to continuously develop a Visual Analytics system's underlying model. In contrast, traditional Business Intelligence focuses on the visualization of data and does not store expert feedback in dedicated databases. As a result, Visual Analytics is more adequate to externalize domain knowledge, which is why we will on Visual Analytics in this thesis.

2.1.1 Models, Approaches, and Theories in Visual Analytics

So far, a large number of models, approaches, and theories have been developed to describe human cognitive capabilities in data analysis tasks. A non-exhaustive list is provided below:

- *Gestalt Principles* [59] are laws of perception that describe how humans recognize patterns, group patterns, or simplify complex images.
- *The Theory of Graph Comprehension* [91] describes how to read graphs and which cognitive operations are executed while inspecting graphs.
- *The Process of Knowledge Discovery in Databases* [36] refers to the extraction of potentially useful information from raw data out of databases.
- *The Classical Visualization Pipeline* [15, 16] describes the process of creating visual representations of data in the four steps, data analysis, filtering, mapping, and rendering.
- *The Skill-Rule Knowledge Model* [93] outlines how skill-, rule-, and knowledge-based performance levels are perceived in terms of signals, signs, and symbols.
- *The Sense-Making Loop* [92, 128] is a process framework to describe analysis tasks, such as the search for information or evidence.
- *Visual Analytics Process Models* [115, 57, 119, 5] all provide high- and low-level descriptions of Visual Analyitcs processes, for example how data is processes and made available to users via specific visualization interfaces [119].
- *The Human Cognition Model* [43] refers to the process of information discovery and knowledge building. Here, the computer presents information that humans can perceive and directly interact with to focus their attention [100]. Knowledge can be created via discovering patterns or relations in the presented information.
- The process of *Knowledge Generation in Visual Analytics* [99, 32] describes how knowledge is created through Visual Analytics systems, for example through exploring and verifying instances of available visualization interfaces.
- The model of *Visual interactive Labeling* [12] proposes a guided workflow of how knowledge can be externalized with labels and how created labels can be used to interactively train and improve machine learning models.
- Knowledge Assisted Visualization [37, 123] outlines ways to incorporate implicit and explicit knowledge into visualization interfaces to support decision making.

Within the scope of this thesis, we used the VA model from Keim et al. [57] as starting point for our research. We consider this model adequate because it allows for a systematic exploration of how Visual Analytics systems can be designed and developed, while also considering the role of human experts that create knowledge through analysis tasks.

2.1.2 Knowledge Creation in Visual Analytics

Many models of VA explicitly involve the notion of knowledge to varying extends [37]. Nevertheless, the models of van Wijk [119], focusing on the system perspective and Keim et al. [57], focusing on the human perspective have been widely adopted and enhanced so far and will thus be discussed in detail.

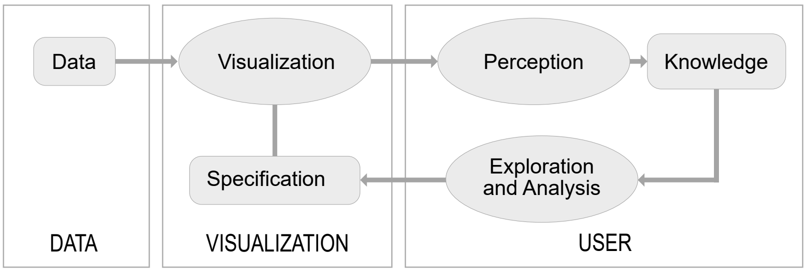


FIGURE 2.2: Knowledege Creation in Visual Analytics according to van Wijk

Figure 2.2 shows van Wijks [119] operational model of visualization. Van Wijk identifies the three spaces, data, visualization, and userspace, to describe the context in which visualizations operate. The model also considers knowledge within the user space in the two dynamic processes: exploration and perception process. The former describes the specifications of algorithms and parameters and the latter the creation of new knowledge through the analysis of visualizations considering prior knowledge [119]. Van Wijks' model has been adapted and extended by various visualization scholars. Green et al. [43] relate their human cognition model for VA to van Wijks' model and propose that exploration, perception, and knowledge should be modeled as cognitive processes to inform each other [43]. [37] extend van Wijks' model and propose a conceptual model of Knowledge-assisted VA, as a high-level abstraction of incorporating explicit and tacit knowledge in the analytical reasoning [37]. However, their model focuses above all on describing system characteristics and does not consider user operations. Finally, Wang et al. [127] extend van Wijks' model by adding a knowledge base to the model and relating it to the process of knowledge creation and conversion from Nanaka and Takeuchi [88].

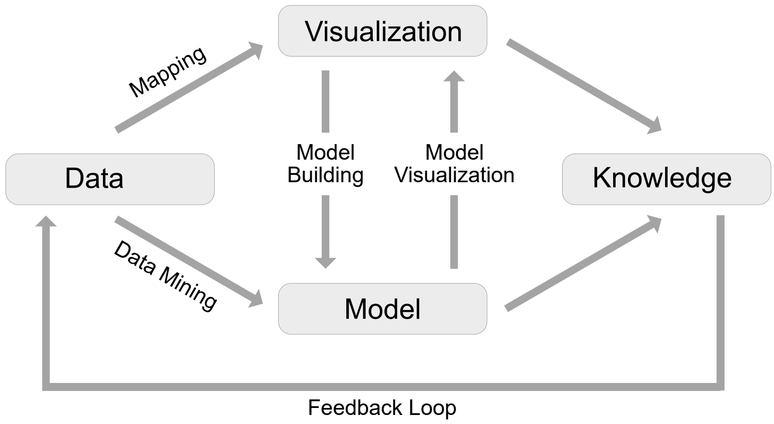


FIGURE 2.3: Visual Analytics process model according to Keim et al. [57]

In comparison to van Wijks' model and its extensions, the VA process model provided by Keim et al. [57] focuses more on user operations, combining automated analysis methods with user interactions to gain insights and knowledge from data and information [57]. Here, users can choose to either apply automated or visual analysis methods. When using an automated method, data mining methods are applied to create models that fit original data and reveal interesting properties of the dataset (e.g., by classifying data points or modeling relationships between the variables comprising the dataset). These models are then used to evaluate and improve the visualization. If the user chooses to conduct a visual analysis, raw data is mapped directly to specific types of visualizations. Alternating between visual and automatic methods leads to the iterative verification and refinement of preliminary results. Findings in the visualization can then be used to steer model building that improves the automatic analysis. Through the interplay between visualizations, models, the underlying data, and the user, new knowledge can be created [57]. Lammarsch et al. [65] extended the VA process model, focusing on including domain knowledge about time-series data from previous analyses. Sacha et al. [100] extend the model of Keim et al. [57], by describing the process of knowledge creation through the interaction of the user with a VA system. The extended model from Sacha et al. [100] is shown in Figure 2.4

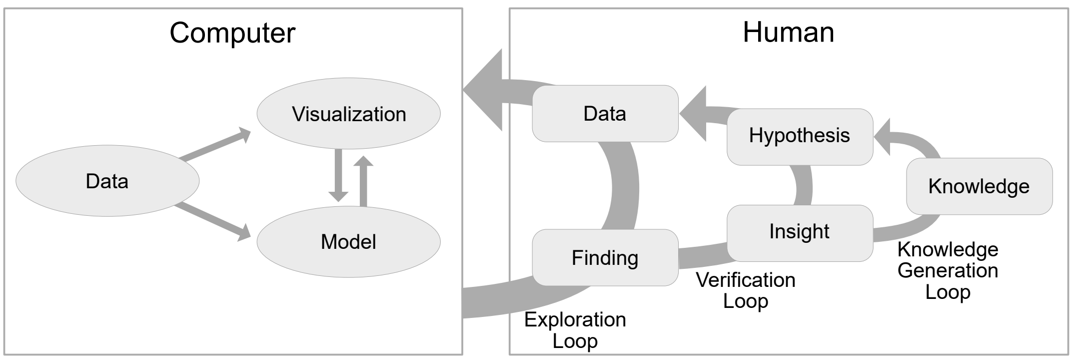


FIGURE 2.4: Knowledge Generation in Visual Analytics according to Sacha et al. [100]

Here, the frequent interaction with the system leads to findings. Findings can either result in actions if they are taken as input for additional interactions with the system, or insights if they are interpreted based on the users' domain knowledge. Insights can lead to hypotheses, that have to be tested properly. To verify a hypothesis a new interaction with the system is triggered by an action. By repeatedly iterating through this process, the VA system affords the creation of new knowledge for the system's user.

In this thesis, we did put close attention on including the user - in our case manufacturing experts - in the design and development of our proposed VA systems. Furthermore, we aim to discover how VA can support the creation, formalization, and transfer of expert knowledge. In this regard, the model from Sacha et al. [100] specifically focuses on the process of knowledge creation through VA, while keeping the user in the loop. However, this model does provide an answer on how to formalize, and transfer knowledge through VA. Hence, we extended the model from Sacha et al. [100] to account for novel mechanisms of knowledge formalization and externalization (See Section 4.2.1).

2.1.3 Knowledge-Assisted Visualization

Apart from modeling knowledge in VA processes, the area of knowledge-assisted visualization specifically aims at including knowledge from analytical tasks of domain experts in VA systems [71, 18, 127, 81]. In our case, analytical tasks included the classification of electrical engines by analyzing their acoustic signatures, the identification, and verification of rules from machine learning models, the provision of labels and annotation for high dimensional sensor data, or the comparison of sensor measurements along with entire manufacturing processes.

Knowledge for such high cognition analytical tasks can be divided into two types: On the one hand, *explicit* knowledge can be "written down" or formalized easily, for example, in words, symbols, or numbers. This process is referred to as externalization [88]. On the other hand, *tacit* knowledge that is inherent to the individual [88, 32, 37, 127]. Tacit knowledge is often not recognized by the individual as knowledge but rather expressed through action, commitment, and involvement, which renders it notoriously difficult to externalize [37].

With the capabilities of generating and externalizing knowledge from knowledge-assisted visualizations, domain experts can perform more efficient and effective data manipulation and analysis tasks [127, 126, 124]. Tacit knowledge from such analyses can be externalized and stored, for example via *rules* [83, 132], *metadata attributes* [85], *graphs* [111, 42, 109, 138, 75], *labels* [97, 49, 131, 11], or *annotations* [94, 28, 31] of data instances.

The resulting explicit knowledge can then be transferred among different users providing a common context that helps to bridge differing backgrounds and to reduce the burden to acquire domain knowledge about different domain problems [81].

Furthermore, different knowledge types provide different support when leveraged in knowledge-assisted visualizations [71, 81, 124]. Here, *operational knowledge* deals with knowing how to use visualization systems, while *domain knowledge* contains knowledge about analyzed data helping users to correctly interpret visualizations. In this thesis, we are particularly interested in how *domain knowledge* can be created, formalized, and transferred via VA.

2.1.4 Applications of Visual Analytics in the Manufacturing Sector

Visualization to date has contributed substantially to help to analyze complex data in manufacturing settings, as the recent survey by Zhou et al. [139] shows. Thus, in this thesis, we were inspired by multiple VA systems designed and developed in the context of industrial manufacturing processes.

Regarding industrial manufacturing processes, VA applications mainly support product design, condition monitoring of stations, the optimization of manufacturing processes, anomaly detection, or the visual support of high cognition tasks. In the following, we provide examples of such applications. For example, Sedlmair et al. [102, 104, 103] developed various VA applications to visualize in-car communication networks. Cibulsiki et al. [19] developed a VA system to facilitate the exploration of multi-criteria alternatives for rotor designs. Suschnigg et al. [114] did build a system to detect and analyze anomalies in test stations, while Xu et al. [133] support the visual exploration of assembling data to detect inefficiencies. Gashi et al. [40] developed a visualization approach to show the interdependencies of manufacturing stations for predictive maintenance. Jekic et al. [54, 55] focused on supporting domain experts in identifying patterns of manufacturing data from aluminum production. Vukovic et al. [122] did build a VA system to forecast important performance quality indicators for sinter production improvement. In the field of anomaly detection, Wörner et al. [130] visualize diagnostic machine data to help identify specific elements in machines, which need to be replaced or repaired. Finally, Maier et al. [76] present a visualization to guide the user in detecting anomalies of time series data in model manufacturing plants.

Throughout this thesis, we were guided by previously developed solutions to tackle industrial manufacturing problems. For example, Suschnigg et al. [114] used methods of anomaly detection to identify anomalous petroleum engines in test benches. In their approach, they visualized an anomalous engine together with an engine without an anomaly. In our system *VIMA*, we also applied this baseline comparison method. However, *VIMA* goes one step further and does not only allow the comparison of two engines but also shows the deviation of a selected engine to measurements of all other analyzed engines.

Another example is the work of Sedlmair et al. [104], who focus on the visualization of the in-car-communication network. As well as all VA systems of this thesis, Sedlmair et al. [104] focused on the role of expert knowledge in their visualization approaches. In this regard, they stored externalized expert knowledge in the form of state machine diagrams, which inspired us during the development of our system *IRVINE* [28]. However, *IRVINE* was built for the analysis of high dimensional acoustic data and hence stores expert knowledge in the form of labels for electric engines and annotations in the raw sensor data instead of using state machine diagrams.

Furthermore, many of the mentioned studies succeeded in creating insights for engineering experts based on machine sensor data. However, they did not consider how users with a much lower level of engineering knowledge, such as data scientists can leverage expert knowledge. To bridge this knowledge gap, we developed the system *ManKnowVis* a system that processes, combines, and contextualizes different knowledge repositories, which are enriched with tacit knowledge from multiple domain experts. The resulting combined knowledge is available for users with less domain knowledge through visual interfaces, which supports them in better-comprehending machine sensor data.

2.2 Knowledge Management

Knowledge Management focuses on the creation and transfer of knowledge inside corporations [24, 26, 53, 69]. In this regard, companies are investing large resources in knowledge management to improve their operations, enhance decision making, or increase productivity [25].

As an integral part of knowledge management, Nohria and Tierney [48] identified two strategies for administering knowledge inside corporations, which are "codification" and "personalization".

Codification can be strongly related to the externalization of knowledge [89], which is the formalization of knowledge with words, numbers, or specifications and the storage in company-wide accessible databases [45]. With the advent of knowledge externalization technologies, such as data mining or wikis, companies are more and more able to leverage dedicated knowledge management technologies to improve their operations [20].

Personalization relates to the tacit knowledge of individual members of organizations, which is often shared through personal interaction, where intuition and social skills play an important role to create and transfer knowledge [20, 88]. To increase knowledge creation and transfer, companies seek to enhance interaction and sharing of tacit knowledge among employees [79], for example through the mediation of learning or shared work practices. This kind of knowledge often is available in for form of highly developed and specific expertise, which can be used to deal with high cognition tasks [45], such as an in-depth analysis of acoustic data from electrical engines.

However, regarding both the codification and personalization of knowledge, little attention has been paid to the role of Visual Analytics. We build on this research gap and demonstrate, how dedicated VA systems can support the creation, externalization, and transfer of knowledge in organizations.

2.2.1 Models, Approaches, and Theories in Knowledge Management

So far, a large number of models, approaches, and theories have been developed to describe knowledge creation, externalization, and transfer. A non-exhaustive list is provided below:

- The *Knowledge Pyramid* [10, 95] is a representation of the structural and functional relationships between data, information, knowledge, and finally wisdom. Here, data is transformed into information via data processing routines. Only if interpreted correctly the information can potentially become knowledge. An example is the knowledge that an anomaly in manufacturing data always can be related to a specific error of a produced engine. In the same example, the knowledge is transformed into wisdom, when the owner of the knowledge also knows *why* the error can be related to the engine. For instance, the reason for the error in the engine can be a broken rotor shaft.

- *Organizational Learning* [8, 7, 56] is the process of creating, transferring, and retaining knowledge within organizations. In this regard, organizations learn and improve over time by gaining new experiences. From this experience, they can create new knowledge, which can be transferred and retained among organizational members.

- The *SECI Model* [88, 89] describes the processes of knowledge creation and conversion in the four phases *socialization*, *externalization*, *combination*, and *internalization*.

- The *LOFT Knowledge Cycle* [137] consists of the four phases *listening*, *observing*, and *feeling* of *thoughts* to achieve stillness, and self-mastery.

- The *Maturation Knowledge Cycle* [3] considers the three knowledge types: *descriptive* (know-what), *procedural* (know-how), and *reasoning* (know-why). Descriptive knowledge describes the state of a domain, procedural knowledge comprises logical procedures and routines, while reasoning knowledge refers to logical methods and skills to solve problems [45].

- The *Wiig Knowledge Cycle* [21] outlines four phases of how knowledge is created and used within organizations. The four phases are *building*, *holding*, *pooling*, and *applying* knowledge. In the first phase (building), knowledge is created and acquired by individuals. The second phase (holding), refers to storing knowledge in repositories or databases. In the third phase (pooling), knowledge is made available to organizational members. In the last phase (applying), the knowledge is used to solve problems, perform analytical tasks, or decision-making.

- The *Maruta Knowledge Creation Mental Model* [78] is a mental model to show how an individual acquires and creates knowledge. It consists of a *memory* and a *processing* function. In the memory function, an individual holds either explicit, or tacit knowledge, and/or other information. In the processing function, the sub-function *insight for comprehension* has the purpose to analyze information, while the sub-function *insight for creation* has the purpose to create knowledge.

The SECI model [89, 88, 87] and the model of Organizational Learning [56, 7] are by far the most discussed models to describe knowledge creation, externalization, and transfer [45]. Hence, we used them as a basis for our research to conceptualize how VA can be used to create, externalized, and transfer knowledge within organizations.

2.2.2 The SECI Model

According to the knowledge-based view, knowledge creation capabilities are a strategic asset that contributes to organizations by helping to improve organizational performance [41]. Nonaka and Takeuchi [88] contributed to the area of knowledge creation with their theory of knowledge creation and proposed their SECI model to conceptualize the process of knowledge creation and the conversion between tacit and explicit knowledge. It is displayed in Figure 2.5 and comprises the processes of *socialization* (S), *externalization* (E), *combination* (C), and *internalization* (I).

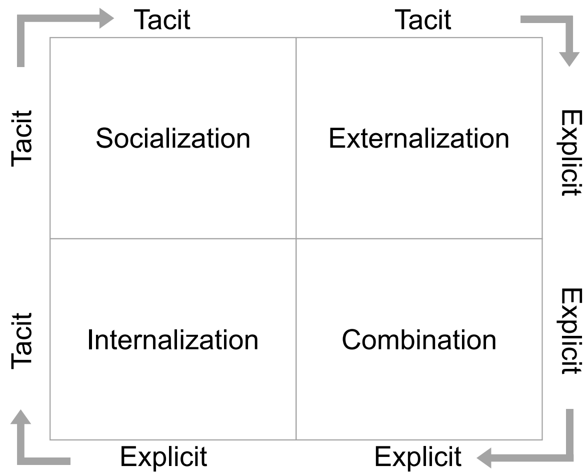


FIGURE 2.5: Process of organizational knowledge creation and conversion according to Nonaka and Takeuchi [88].

During interactions between individuals (socialization), tacit knowledge is converted through shared mental models. Resulting tacit knowledge is converted into explicit knowledge by formalizing it (e.g., in the form of words; externalization). Externalized sources of explicit knowledge can be combined into more systematic and comprehensive sets of explicit knowledge (combination). Further, explicit knowledge can be transformed into tacit knowledge by individuals, which is closely related to the concept of "learning by doing" (internalization) [89].

The sequential iteration of each process phase of the SECI model creates a spiral of new knowledge. To increase knowledge, a central task of organizations is to ensure knowledge creation and conversion [52].

Prior research supports the notion that the use of information technologies is positively related to the collection, storage, and dissemination of knowledge in organizations [58, 7, 52, 64, 116]. This previous work, however, focuses on storing and sharing information [58, 64, 116], neglecting the role of model-based reasoning or interactive data visualization in knowledge creation in collaborative contexts, as they are afforded by VA systems [86]. We build on this prior research to include these additional dimensions and analyze their effects on knowledge creation between VA users and their broader environment.

2.2.3 Organizational Learning

Organizational learning focuses on the dynamic processes through which knowledge is created and consumed within organizations [121]. It can broadly be divided into two forms: exploitation and exploration [56]. The former refers to incremental learning, which focuses on the diffusion, reuse, and refinement of existing knowledge [66, 77, 108], whereas the latter involves the development of new or the replacement of existing knowledge in organizations [1, 77, 90]. In line with the notion of exploration and exploitation of knowledge in organizations, Argote and Miron-Spektor [7] provide a conceptualization of organizational learning (see Figure 2.6). The key elements comprise task performance and experience, knowledge, and the active context; the latter depends on the latent organizational context. The framework provides an analytical representation of learning in organizations in a cyclic relationship between the three key elements, in which task performance and experience create or transform knowledge through interaction with the context. While the environmental context represents elements outside an organization (e.g., suppliers or customers), the latent context represents elements within the organization (e.g., a culture of trust in technologies). Although the latent context does not "act" actively, it affects learning through its influence on the active context. In particular, within the active context organizational members (e.g., employees) and tools (e.g., Visual Analytics systems) perform actions, that is, tasks related to their jobs.

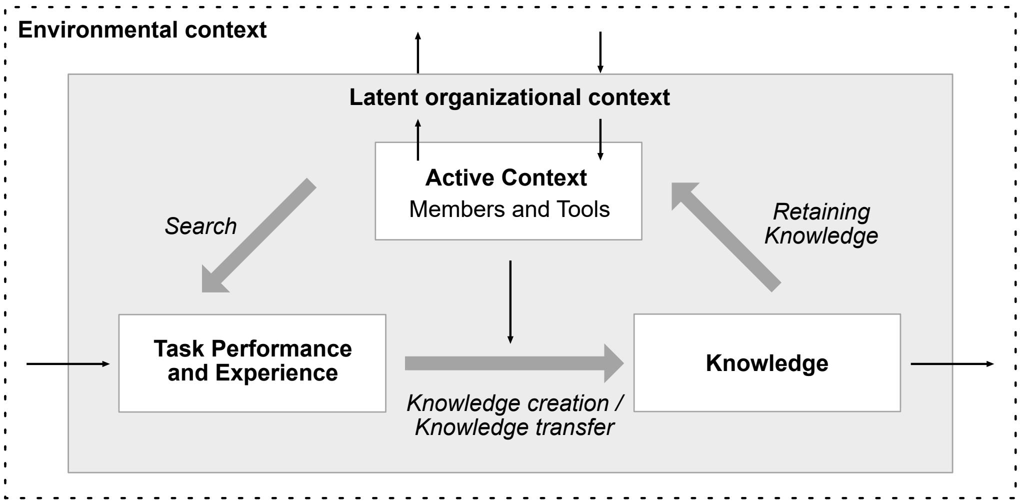


FIGURE 2.6: Process of organizational learning according to Argote and Miron-Spektor [7]

The process in which the active context affects task performance and experience is defined by Argote and Miron-Spektor as search. Knowledge is acquired through new experience (knowledge creation) or the sharing of existing experience within the organization (knowledge transfer) (Kane et al., 2005). This knowledge can either be of tacit or explicit nature [88]. While tacit knowledge is bound to the individual and difficult to communicate, explicit knowledge can be readily accessed by members or tools. Knowledge is retained through the active context, where organizational members or tools use it to act. Organizational members and tools, hence, represent knowledge repositories [2] and generate new experiences by performing new tasks.

Research on the use of data-driven approaches, such as Visual Analytics in organizational learning, stresses that data-driven approaches shift the focus from the human learner to the machine. In contrast to traditional knowledge management tools (e.g., expert systems), novel technologies, such as Visual Analytics not only support human learning but can learn autonomously or learn in form of a hybrid practice where humans and algorithms inform each other [112, 118]. In such cases, human knowledge is crucial to inform human-in-the-loop practices, where domain experts provide inputs for algorithms, training or debugging models, and for making sense of the results. These interactive learning approaches not only allow the integration of human domain knowledge in the design of the system but require a huge amount of formalized data provided by human experts, for example in the form of labels. However, human-generated formalized knowledge can be error-prone. In knowledge-intensive domains characterized by uncertain or ambiguous knowledge work such as in generating diagnosis outputs, formalized knowledge may rely on the opinion of a single person and lack external validation. Regarding knowledge externalization tasks in such domains, the difficulty of identifying false negatives may be exacerbated because the outcome of the diagnosis may only be validated long-term, taking months or years [67].

Hence, especially in situations where an objective ground truth is missing, it is crucial to diligently monitor and assess the quality of externalized knowledge and the system output, (e.g., a suggestion of a relevant class from a Visual Analytics system) to avoid damage due to wrong decisions based on incorrect system suggestions.

In fact, a key challenge of a hybrid human-machine learning process that builds on feedback is identifying and understanding the gap between the actual and reference output. Any data-driven modeling approach, hence, requires human auditing when being built, which, in turn, requires a reference measure (i.e., the ground truth provided by the domain expert) against which the output is compared. Therefore, data analysts and domain experts need to closely collaborate to alter and audit the collaboratively build system and improve its performance over time [46]. While research in this area focused on the development and adoption of data-driven approaches for organizational learning and its intricacies [46, 118], we know little about the role of formalizing and sharing knowledge in this process. In response to this gap, this thesis also investigates how formalized expert knowledge influence organizational learning in the context of industrial manufacturing processes, revealing challenges in building and use of labels as well as the effects of formalized knowledge and related tools, such as Visual Analytics system on the key elements of organizational learning.

Chapter 3

Research Methodology

Several design frameworks and models, as well as different evaluation methods, contributed to the development of the research work presented in this thesis. The current research aims to evaluate how Visual Analytics can support the creation, externalization, and transfer of domain knowledge in industrial manufacturing setups. In this chapter, we succinctly present frameworks, models, and methods, we used to carry out our research projects. Furthermore, we discuss the overall research procedure and the main activities conducted in the context of this thesis.

3.1 Design Study Methodology

The most important methodology used in this thesis is the design study methodology according to Sedlmair et al. [105]. Design studies are an approach to performing problem-driven research, where the focus lies on collaborating with real users to address real-world problems. Furthermore, design studies are a form of technique-driven research to apply and further develop established techniques to solve user problems. Hence, the contribution of design studies goes beyond "just" developing practical. In fact, they help in abstracting and contextualizing real-world problems and in developing general principles to address such abstract problems with visualization techniques. That said, in this thesis, we rely on the definition of design studies from Sedlmair et al. [105] - page 2.

"A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect on lessons learned in order to refine visualization design guidelines."

To evaluate whether a design study is a suitable research methodology, Seldmair et al. [105] propose to analyze a real-world problem along with the two axes of task clarity and information location, as shown in Figure 3.1.

The **task clarity** axis depicts how precisely a task is defined from fuzzy to crisp. An example of a crisp task is to "cook a cup of tea". Here, the task has a clearly defined goal and is addressed with a known set of straightforward steps, such as "boiling water" and "putting a tea bag into a cup". For such crisp tasks, it is hence relatively easy to design and evaluate solutions. However, reducing real-world problems to these tasks is challenging, where researchers are often confronted with more complex and fuzzy domain tasks. Examples of this thesis are the analysis of high dimensional acoustic signature of electrical engines [28], to understand the decision-making process of a random forest [35], or to identify root cause errors along an entire manufacturing process [33]. These tasks are not well defined and of exploratory nature.

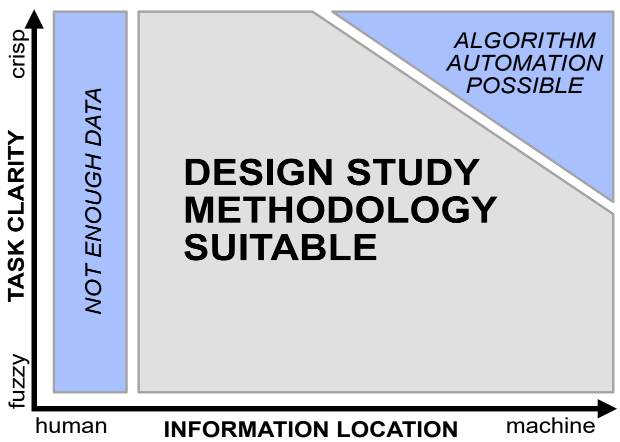


FIGURE 3.1: The task clarify and information location axes to evaluate the suitability of design study methodology according to Sedlmair et al. [105]

As a result, Sedlmair et al. [105] propose to decompose high-level tasks into a narrow and low-level set of abstract tasks. In one of our studies, examples for the analysis of acoustic signature from electrical engines [28] are to *Assign a label to an engine* and to *annotate the acoustic measurement of an engine*. These task abstractions not only help to specifically design solutions (e.g., custom build labeling and annotation interfaces) for individual tasks but also help researchers to address similar real-world problems. For example, in the case of our system IRVINE, acoustic signatures of electrical engines contain a similar data structure to sonograms. Sonograms, however are a common representation of speech signals [38], seismic activities [17], or medical images [51]. As a result, our derived high-level tasks for labeling and annotating could also be used to design similar Visual Analytics systems to formalize expert knowledge in these domains.

The second axis in Figure 3.3 depicts the **information location**. This axis shows how much information is available in the "head" of a human expert versus what has already been made explicit to the "machine". In other words, this axis depicts how much information and knowledge remains as tacit knowledge [88] in the human expert, versus how much knowledge is already formalized that can be used in the visualization.

Along these two axes, researchers can evaluate if a design study is a suitable choice. Figure 3.3 shows how design studies fall along the two axes. The blue areas represent situations, where design studies might not be the best methodological choice.

The first situation is when **not enough data** is available. In this area, effective visualization design is not likely to be possible. The other situation is when an **algorithm automation is possible**. Here, visualization might be inappropriate because already enough information is formalized to develop fully automatic solutions. Since many real-world data analysis problems have not yet advanced to a crisp task where all necessary information is already interpretable by machines, Sedlmair et al. [105] propose that design studies are also a step towards a final goal of fully automated solutions.

The remaining gray area indicates situations where design studies are appropriate. This area is largely resulting in different design studies with different characteristics. In Figure 3.2, we outline four exemplary scenarios, where a design study is a suitable research approach. In (1), the task may be clear but the data could be located in the head of an expert. Hence, this study would require a significant effort in the data characterization. In turn, in the second scenario (2), the task may be fuzzy and only little data be already formalized. Thus, apart from a focus on data characterization, this type of design study would also need a good abstraction of the domain problem and thus the resulting tasks. In the third scenario (3), the task could be crisp, and a lot of data is already accessible to machines. Here, the focus should lie more on the visual encoding and on the technical contributions on how to address well-defined tasks with novel visualization techniques. In this scenario, the goal is above all to reach a higher level of automation. In the fourth example (4), a lot of information could already be accessible in a machine-readable format. However, the task would still be fuzzy. One approach here could be to visualize model behavior and predictions to achieve a more crisp task abstraction.

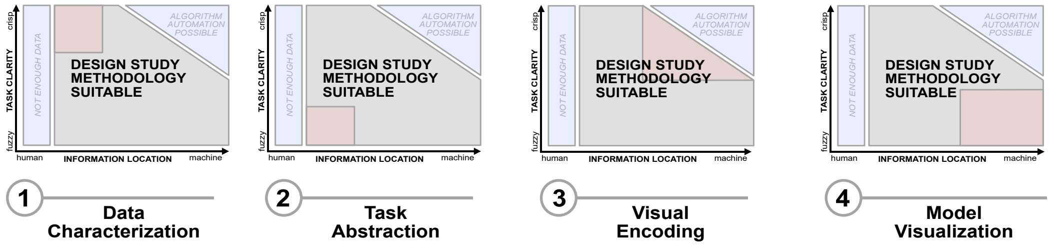


FIGURE 3.2: Four exemplary scenarios to apply design studies

3.1.1 Nine-Stage Framework to Perform Design Studies

When the situation is suitable to perform a design study, Sedlmair et al. [105] also provide a nine-stage framework to perform design studies. The original Design Study Framework is shown in Figure 3.3, which is divided into the following three phases: The **precondition** phase describes what must be done before starting a design study. The **core** phase presents concrete steps to perform a design study. Finally, the **analysis** phase depicts analytical reasoning processes at the end of the study. Each of the three phases contains also a validation stage, which is different for each phase. In the preconditioning phase, the validation occurs on a **personal** level. Here, the validation depends on the personal preparation and experience of the researcher for the project, including due diligence before committing to a collaboration with real users and other collaborating researchers. In the core phase, validation is **inward-facing**. This validation type emphasizes evaluating findings and artifacts with domain experts, for example in the form of interviews or workshops. In the analysis phase, the validation is **outward-facing**. In this validation type, the focus is on the justification of the study results to the outside world.

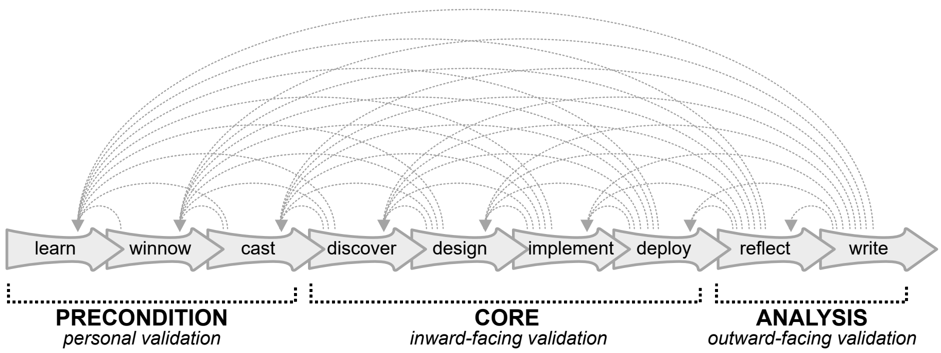


FIGURE 3.3: Nine-stage design study methodology according to Sedlmair et al. [105].

Each of the three phases comprises subphases, which we will describe in the following:

Learn: The first precondition for conducting a design study is a solid knowledge of the respective literature for the research topic, for example about specific visual encodings, statistical methods, or knowledge generation mechanisms. This literature knowledge informs all later stages. For example, the winnow phase supports the selection of research collaborators with respective knowledge about the research area or skill sets to support the design of Visual Analytics systems. Another example is the design phase, where discovered literature helps to broaden the consideration space of possible solutions and to compare newly developed solutions with already existing ones.

Winnow: Probably the most important aspect of research is to collaborate with other researchers and real-world parties, such as industrial domain experts. Hence, the goal of the winnow phase is to identify the most promising collaborators to perform the design study. In Section 1.3, we already outlined a broad list of collaborators, who strongly supported the success of this thesis and the broad list of our publications in different research areas. For example, we strongly collaborated with Michael Sedlmair to support us in performing design studies, with Jürgen Bernard as an expert in interactive labeling, Diana Fischer-Preßler with much experience in the field of information systems, or Jakob Bonart with lots of expertise in the field of the acoustic behavior of electrical engines.

Cast: The cast phase is about relevant roles during a design study collaboration, where one can distinguish between the two main roles. The **front-line analyst** and the **gatekeeper**. The **front-line analyst** is the end-user and domain expert doing the actual data analysis tasks and uses the developed visualization tool. The **gatekeeper** is the person, who approves or blocks the project, including the authorization of other involved people to spend time on the project or realize relevant data.

Discover: The first sub-phase in the core phase is to characterize and abstract the domain problem. In this sub-phase, it is crucial to learn about the target domain, the practices, needs, problems, and requirements of domain experts to discover how information visualization can enable new insights and discovery. This phase is closely related to requirements analysis from the field of software engineering [62], which also focuses on a close collaboration with domain experts. The process of abstracting and characterizing the domain problem is an iterative process, where experts talk to researchers, who then have to abstract and present the abstractions to the researcher. While the abstraction begins in the discover phase, it continues throughout all stages of the nine-stage framework. Abstractions heavily support the transferability of the design study results to other domains. In this thesis, we provide several abstractions that inform other related research areas, which are on the one hand side outlined in our publications and on the other side summarized in Section 5.2.3.

Design: After the characterization of the problem and abstractions of relevant tasks, in the design sub-phase the actual visualization interface is designed. This phase is above all generating and validating data abstractions, visual encodings, and interaction mechanisms. The design phase also includes the selection of proper algorithms and experimentation with data processing routines to provide a foundation to successfully implement visualization interfaces.

Implement: The implementation of the visualization system is tightly connected with the design process. Here, the design decisions are implemented and tested with users, for example via rapid prototyping. As well as in other phases, in the implementation phase, it is crucial to test implemented design alternatives with users and to critically discuss different alternatives for the final visualization interface.

Deploy: In the last sub-phase of the *core* phase, the visualization interface is deployed to gather additional feedback and to use it *in the wild*. For example, in our design studies [30, 35, 28, 33, 31], we designed and implemented each Visual Analytics system in close collaboration with few domain experts and tested it with at least four different domain experts after the system's deployment. To evaluate our visualization approaches, we used several methods, which we describe in detail in Section 3.2.3.

Reflect: The first sub-phase of the *analysis* phase is to reflect on the performed study and its value for research. Through reflection, research emerges from engineering, where the outputs of a design study can be related to a larger research area, especially in the field of information visualization. The reflection is particularly helpful for improving existing visualization design guidelines. For example, based on findings, guidelines can be confirmed, refined, rejected, or even new guidelines can be proposed. In this thesis, we also proposed a variety of new guidelines, which are outlined in each design study publication. These findings are furthermore summarized in Section 5.2.2.

Write: The writing about a design study can be started at any point of the study. The writing of a publication about the design study is the perfect time to revisit abstractions, identifying contributions, and come up with a coherent and understandable line of argumentation.

3.2 Additional Research Methods

With the description of the design study methodology at hand, we can describe the general research methodology for this thesis in Figure 3.4. Each research project is motivated by a real-world use case from an industrial manufacturing setting. One example was that it was hard to compare data from multiple manufacturing steps, which resulted in the Visual Analytics system ManEx [33]. Each use case was addressed with a design study resulting in a Visual Analytics system. With this approach, we developed five Visual Analytics systems in total. Each Visual Analytics system was properly evaluated (See Section 3.2.3). The Evaluation of the systems resulted in two scenarios.

First, we used the findings from each design study to draw theoretical implications, which we address in separate research projects. For example, the findings from the Visual Analytics system VIMA [30] resulted in a separate case study to discover novel mechanisms for organizational knowledge creation through Visual Analytics [32].

Second, we used the findings from a design study as practical implications to better address future use cases. For example, during the development of the system IRVINE [28], we discovered a novel method to represent anomalies as radial glyphs. As a result, we used the same representation in the following design study to visualize anomalies in the manufacturing of electrical engines with the system ManEx [33].

Apart from design study methodology, we also heavily relied on other research approaches, for example, the visualization mantra from Schneiderman [106] *"Overview first, details on demand"*, which we briefly describe in this section.

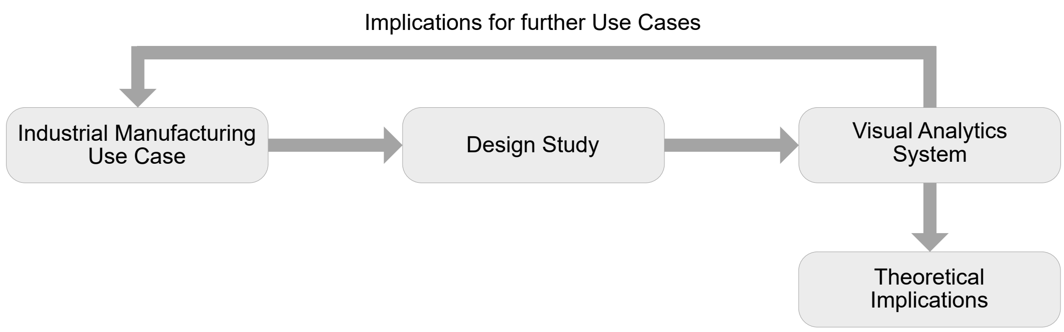


FIGURE 3.4: High-level research procedure of this thesis.

3.2.1 The Design Triangle

One model to guide the design of Visual Analytics systems is the design triangle from Miksch and Aigner [80], which focuses on the **users**, **data**, and **tasks**. Figure 3.5 shows the design triangle. Miksch and Aigner [80] argue that the consideration of **data**, **users**, and **tasks** determines which visual representation, interaction methods, or analytical means are suitable for given analysis scenarios. In addition to that, they complement this design methodology with three qualitative requirements to guide design choices, which are *expressiveness*, *effectiveness*, and *appropriateness*. *Expressiveness* refers to the ability of the visualizations to show the exact information contained in the data [74]. The next criterion *effectiveness* considers the degree to which visualizations address the cognitive capabilities of humans, as well as the analysis task at hand or the application background, [74]. The last criterion *appropriateness* involves the consideration of a cost-value ratio to assess the benefits of the visualization system to achieve a given task [125, 120].

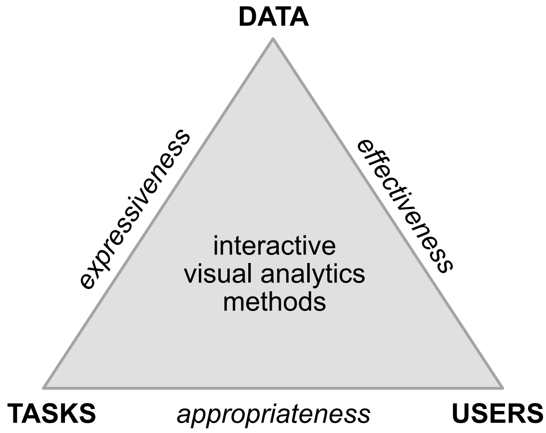


FIGURE 3.5: Design-triangle by Aigner and Miksch [80].

3.2.2 The Nested Model for Visualization Design

Another user-centered design methodology we relied on during this thesis is the nested model by Tamara Munzer [84], which is shown in Figure 3.6. This methodology describes how to design and implement visual representations of data starting from the analysis of the domain problem until the algorithm design. The first phase aims to structure and characterize the problem. The second phase maps the design to the problem and characterizes data types and operations. The third phase is about choosing the right visual encoding and interaction techniques to address previously defined abstractions. In the last phase, the algorithms to implement the selected visual representations and interaction techniques are selected. This framework complements very well the design study methodology [105]. Furthermore, it not only supports the design of Visual Analytics systems but also allows the constant evaluation of the outcomes of each of the four phases.

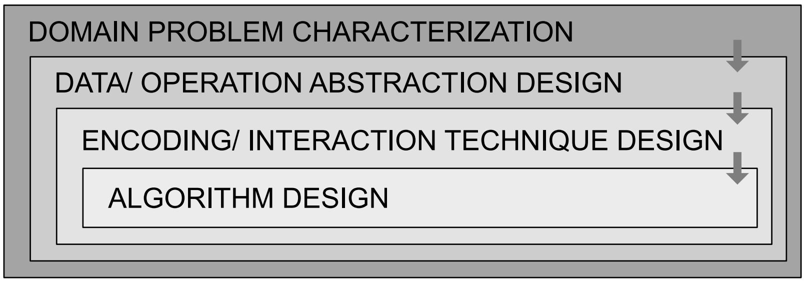


FIGURE 3.6: The nested model for visualization design and evaluation by Tamara Munzer [84].

3.2.3 Evaluation Methods

One critical step during the design and development of Visual Analytics systems is their constant evaluation. Evaluations in Visual Analytics are typically performed in an iterative fashion to continuously improve the designed visualization interfaces to identify problems as early as possible. This approach often resulted in significant changes in our visualization design. For example, in the system ManKnowVis [31], we started with a forced layout representation of an assembly line for battery modules, which we then changed to a linear representation of individual manufacturing steps, and finally to multiple linked views to represent separate manufacturing instances, such as product structures, manufacturing steps, or manufacturing stations.

In the early stages of each system design phase, we mostly relied on semi-structured interviews or workshops with domain experts, critical discussions with our researchers and external collaborators, and our own previous experiences from already performed design studies. This helped us above all to abstract the problem domain, characterize the data, and define specific tasks and requirements that supported us in designing each Visual Analytics system. At the end of each design study project, we further evaluated quantitatively how each system impacted real-world manufacturing settings. For example, in the system IRVINE [28], we compared how much labeling speed improved after the system introduction. In turn, for the system RfX [35], we analyzed how much testing time would be reduced when using the system.

Apart from that, we relied on three evaluation methods, which we briefly describe in the following.

System Usablity Scale: To quantitatively assess the usability of each Visual Analytics system, we applied the System Usability Scale [68]. This scale is composed of ten statements rated on a Likert scale.

Think-Aloud Method: In this research method the system user is asked to use the Visual Analytics system by performing previously defined tasks, for example *"identify an anomaly from the manufacturing process"*. While performing the task, study participants are asked by a present researcher to express their thought out loud [110].

Case Study Methodology: A case study is an empirical research approach to analyzing a specific phenomenon within its environment, especially when the boundaries between the practical context and the phenomenon are not clearly evident [136]. It is useful if the research is not well developed and particularly where the examination of context and dynamics are important [22]. Case studies, furthermore, can be applied to investigate casual relationships and are a suitable instrument for studying context-rich sociotechnical systems [136].

Chapter 4

Results

This chapter describes the results of this thesis in two parts. First, we provide a summary of each design study project, which resulted in a fully running Visual Analytics system. This is followed, by a description of the theoretical frameworks, which were published during the time of the thesis.

4.1 Visual Analytics Systems

As one core part of this thesis, we provide an overview of all five Visual Analytics systems, we developed. Each subsection is structured as follows. First, we provide an overview of the content of each individual Visual Analytics system. Next, we provide insights into each research approach of each study. Next, we summarize the results of each study and provide information on how each Visual Analytics system proved to be an inspiration for future Visual Analytics systems as already outlined in Figure 3.4. Last, I report on scientific contributions and written content in each publication of myself. For details of each publication please refer to Chapter 6.

4.1.1 Publication 1: Virtual Interactive Manufacturing Assistant

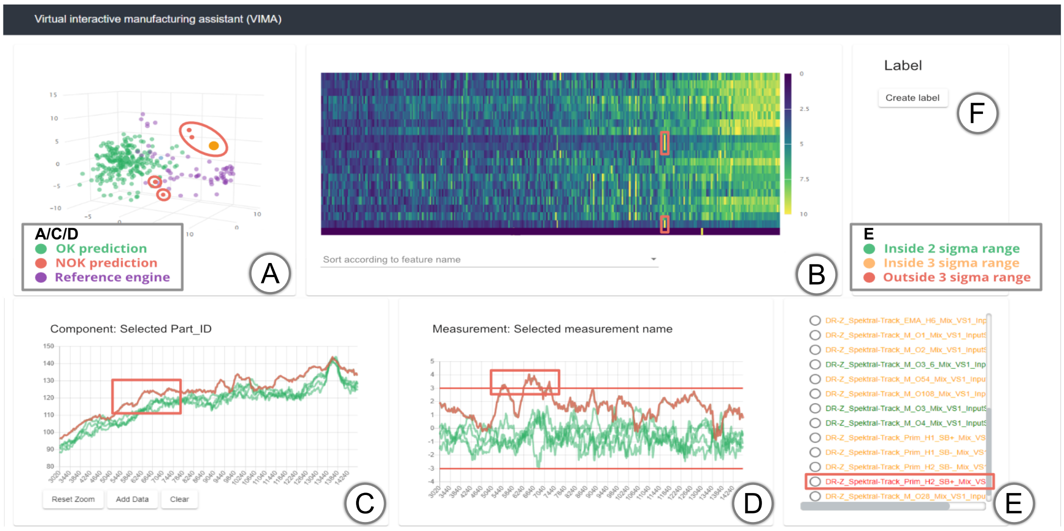


FIGURE 4.1: Screenshot of VIMA. The two views at the top (A, B) provide an overview of the behavior of produced engines, where (A) is the projection of features in 3D via principal components; (B) is a sorted heatmap of all normalized features according to the first component. The views at the bottom (C, D) represent the real measurement (C) and its normalized residuals (D), with corresponding colors to (A); (E) provides all available measurements for a selected engine; (F) provides the user with the possibility to label the selected part.

Citation: *J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In Visualization in Data Science at IEEE VIS, 09 2020*

Content: The highly integrated design of electrical vehicles results in new challenges for in-line quality testing of manufacturing processes. Even though the technology of electrical vehicles is not novel as such, it still remains a challenge to ramp up automated manufacturing processes to produce such complex products. In this context, automated quality testing is firmly integrated into the manufacturing process. Engineering experts, however, face the challenge that automated testing procedures for electrical vehicles are about to be developed. In this regard, such experts more and more rely on sophisticated data analyses of IoT sensor data in automated part testing. Due to large amounts of sensor data, however, engineering experts require automated support in data analysis tasks. In this design study, we present the *Virtual Interactive Manufacturing Assistant (VIMA)*, a Visual Analytics system, which processes high dimensional sensor data to support engineering experts with their analysis of produced parts in the power train for electrical vehicles. VIMA hereby computes residual values of the sensor data to better detect parts with anomalous sensor measurements. It then projects the high dimensional sensor data into a three-dimensional space to better detect statistical outliers, which can be selected for an in-depth analysis. After an analysis is complete, engineers can manually assign a label to the selected part, which can then be used to train a random forest to suggest new relevant engines for labeling.

Research Method: The study was performed in the form of a design study, where we developed the system in close collaboration with two engineering experts from

BMW, one responsible for testing procedures for high voltage batteries and the other for acoustic testing of electrical engines. In the study, we particularly relied on the design triangle by Miksch and Aigner [80]. Regarding the **users**, our system targets engineering experts in the manufacturing of electrical engines. As **data**, we analyzed multivariate time-series data from measurements generated at test benches from manufacturing processes for electrical vehicles at BMW. Regarding the tasks, we abstracted the four tasks (1) *Identify Abnormal Part*, (2) *Analyze Abnormal Measurement, (3) Compare Multiple Produced Parts, and (4) Interactively Label a Produced Part.* To qualitatively evaluate the system, we performed a user study with four engineering experts from the domain of acoustic testing for electrical engines. In this context, we uploaded sensor data and gave the users the task *"please identify an abnormal engine according to your point of view"*. After that, each participant was asked to perform the task following the think-aloud method [110], followed by open-ended questions to evaluate how every single view supported the task.

Results and Implications for Future Studies: With the help of VIMA, engineering experts were able to analyze electrical engines more efficiently and much faster. Furthermore, during the usage of VIMA labels were created, which we used to train a random forest to predict the detect *increased backlashes* as one common error type of the analyzed test bench. Our results show that the random forest outperformed the current testing procedure to detect *increased backlashes* and improved the test benches output by 15%. However, VIMA did not include a functionality to automatically label parts and retrain the random forest. Thus, we took this implication and further developed the idea of interactive labeling in the system IRVINE [28]. Furthermore, users reported that they wanted to know why the random forest performed certain decisions. Hence, we conducted a follow-up design study and created the system RfX [35] to visualize the decision-making process of a random forest.

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Extended state of the art data processing routines of machine sensor measurements to detect anomalies in electrical engines with residual values.
- Conceptualized VIMA's underlying models to support users' abilities to detect anomalous electrical engines. For example, the computation of a principal component analysis to project electrical engines into a three-dimensional space or a random forest to predict the status of an engine.
- Conceptualized interaction techniques, for example, the use of the colors green, orange, and red to visualize anomalies.
- Designed and developed all interfaces.
- Evaluated the system with four engineering experts with knowledge about the acoustic behavior of electrical engines.

Written Contents Contributed by Myself: Besides Section 2.3 about interactive labeling, I wrote the entire contents of the paper, including all tables and figures.

4.1.2 Publication 2: Random Forest Explorer

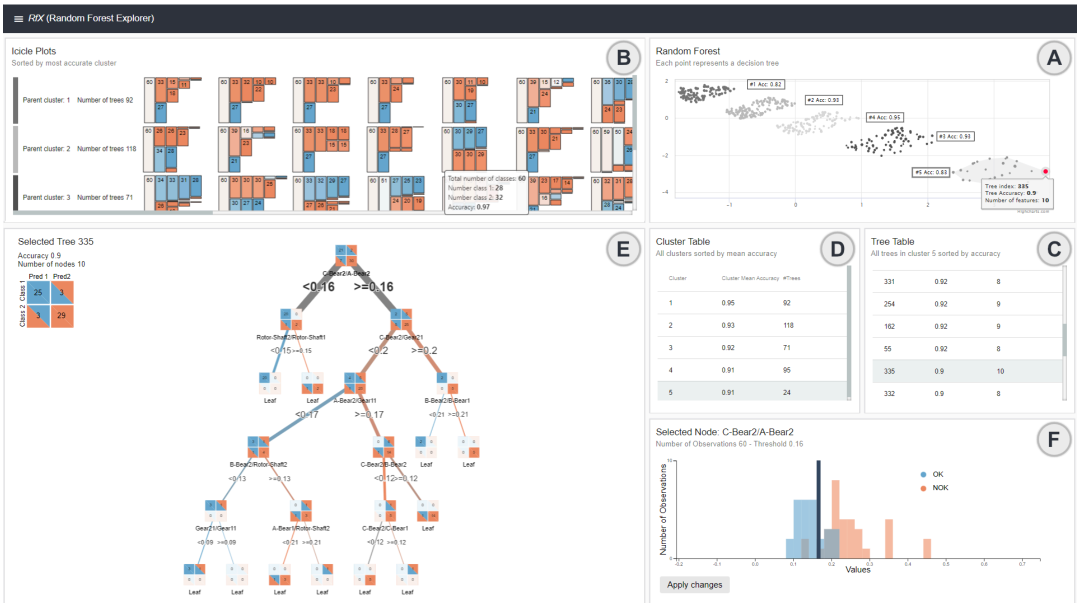


FIGURE 4.2: Screenshot of *RfX*. *(A)* visualizes the projection of all decision trees of the random forest into a two dimensional space. *(B)* shows most representative tress of each cluster, represented as an icicle plot. Each cluster of similar trees is additionally visualized with a table representation, where *(D)* shows each cluster and *(C)* each tree within a cluster. A decision tree and its properties are shown as a node-link diagram in *(E)* and the class distribution of a feature as a histogram in *(F)*.

Citation: *J. Eirich, M. Münch, D. Jäckle, M. Sedlmair, J. Bonart, and T. Schreck. Rfx: A design study for the interactive exploration of a random forest to enhance testing procedures for electrical engines. Computer Graphics Forum, pp. 1–14, 03 2022*

Content: Random forests are a well-adopted and widely used machine learning technique across industries. The interpretation of a random forest usually relies on the analysis of statistical values and often is performed by data analysis experts. As feedback from the design study project VIMA [30], we designed and developed the *Random Forest Explorer* (RfX) to make this machine learning model accessible to engineering experts with no background in sophisticated data analysis methods. With RfX, users can interactively inspect the properties of a random forest. They can easily detect, evaluate, and compare representative decision trees for a region in the random forest. In this regard, engineering experts have different entry points into an analysis, for example with a 2D projection, table views, or icicle plots. A selected decision tree can be further evaluated in detail with a node-link diagram view, where the nodes consist of confusion matrices to visualize the performance of each split. As the lowest level of detail, engineering experts can inspect the distribution of measurement recordings in each node and optimize the split threshold to get a better understanding of the random forest's decision-making process.

Research Method: RfX was designed and developed in the form of a design study project [105], where we also used the nested model by Munzner [84]. A formative evaluation was carried out with two engineering experts in the field of acoustic testing for electrical engines from BMW. We furthermore, performed a summative evaluation in the form of a field study with five domain experts. As **data characterization**, we abstracted a random forest as a sample of decision trees, where each tree can be represented as a graph containing nodes and edges. Regarding the **task abstraction**, we discovered the four tasks (1) *Partition Random Forest*, (2) *Identify Individual Representative Decision Tree*, (3) *Explore Individual Representative Decision Tree*, and (4) *Optimize Decision Tree*. To qualitatively evaluate the system, we performed a user study with five engineering experts from the domain of acoustic testing for electrical engines. In this context, we uploaded sensor data and gave the users the task *"please derive a rule from a feature combination that you found in a decision tree"*. After that, each participant was asked to perform the task following the think-aloud method [110], followed by open-ended questions to evaluate how every single view supported the task. To quantitatively assess the usability of RfX we applied the system usability scale [68].

Results and Implications for Future Studies: Our results suggest a high usability and usefulness of IRVINE as part of the improvement of a real-world manufacturing process. The score from the usability scale was 86, which is well above the average score of 68 [68]. With RfX, engineering experts were for the first time at BMW able to comprehend the decision-making process of a random forest. During the analysis of decision trees from the random forest, engineers discovered new hidden patterns in their sensor data, such as increased eccentricities in an engine's rotor can be also detected by observing secondary excitations of its bearings. This resulted in a potential reduction of testing time of over 80% of the analyzed testing station for over 30% of all produced parts. Furthermore, we compared the prediction accuracy to detect rotor errors of derived rule sets from RfX to a traditional random forest, as well as a traditional decision tree. Our results show that the derived simple rule sets from RfX outperformed decision trees and performed worse than a random forest. However, the derived rule sets can be represented as simple if-else conditions, which are far easier to understand than the decision-making process of a random forest without using RfX. Apart from that, we developed novel visualization techniques to represent the decision-making process of a random forest, with icicle plots or a node-link diagram combined with confusion matrices.

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Conceptualized and implemented the system's back-end to automatically process and analyze hundreds of decision trees from a random forest.
- Abstracted the problem domain, the data, and analysis tasks.
- Developed a semantic and syntactic dissimilarity metric to identify the most representative decision trees out of a random forest.
- Conceptualized interaction techniques, for example, the representation of a decision tree as an icicle plot or node-link diagram.
- Designed and developed all interfaces.
- Evaluated RfX with five engineering experts with knowledge about electrical engines.
- Performed an additional evaluation to compare derived rule-sets with RfX against a traditional random forest and decision tree.

Written Contents Contributed by Myself: Besides the equations, I wrote the entire contents of the paper, including all tables and figures.

4.1.3 Publication 3: Interactive Clustering and Labeling

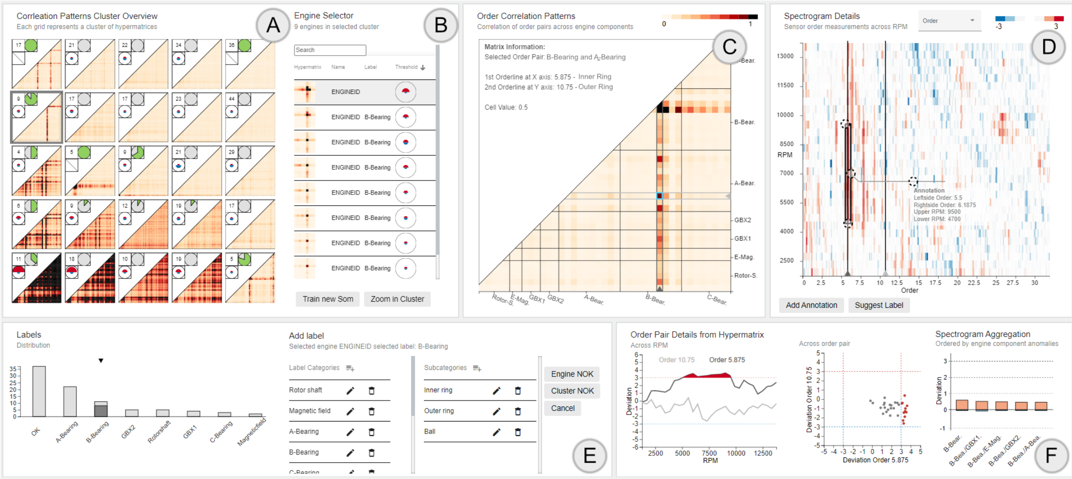


FIGURE 4.3: Screenshot of IRVINE. Users have an overview of clusters in (A). They can select clusters in (A) and engines in (B). After selecting an engine in (B), the acoustic signature of the engine is displayed in (C) and respective raw acoustic measurements in (D). Detailed information about selections from (C) is shown as a line chart and scatter-plot and bar chart in (F). After the analysis of an engine, the user can assign a label in (E) and provide an annotation for the label in (D).

The publication got the award for best paper at the conference IEEE Visualization Conference (The largest and most prestigious conference for information visualization worldwide).

Citation *Eirich, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. IEEE Transactions on Visualization and Computer Graphics, 07 2021*

Content: In serial manufacturing, signatures from acoustic data provide valuable information on how the relationship between multiple produced parts serves to detect and understand primary and secondary error causes. However, detecting and understanding such errors is currently hindered by two challenges. (1) Industrial manufacturing processes produce large quantities of parts, resulting in vast amounts of sensor data. (2) Due to the high amount of produced parts and data, only view engines can be analyzed in detail. Thus, the gained knowledge is particularly precious for the overall improvement of manufacturing processes. Yet, before the study, it was unknown how the gained knowledge can be stored or transferred. In this design study, we address the two named challenges with IRVINE (InteRactiVe cluserINg labEling). We base our study on acoustic data from a test of test benches for electrical engines. In general, acoustic signatures comprise one primary and multiple secondary error symptoms. As part of this study, we developed the concept of the so-called *Hypermatrix*, an abstraction of acoustic raw data to better identify primary and secondary error symptoms. To better analyze both the high dimensional acoustic raw data, as well as the *Hypermatrices*, IRVINE leverages interactive clustering and data labeling techniques to allow users to analyze clusters of engines with similar acoustic signatures, drill down to groups of engines, and select engines of interest. Furthermore, with IRVINE engineering experts can assign labels to engines and clusters and annotate the cause of an error in the acoustic raw measurement of an engine. Labels and annotations represent valuable knowledge of engineering experts and are stored in a knowledge database to be available for other stakeholders.

Research Method: IRVINE was designed and developed in the form of a design study project [105], where we also used the nested model by Munzner [84]. A formative evaluation was carried out with one engineering expert from BMW. First, we abstract engines as the analysis object of the study into the two categories **engines known** and **engines unknown**. The former engines, with errors the engineers are already familiar with, and the latter engines with errors that are new to the engineers. As **data characterization**, we abstracted acoustic data in the form spectrograms to a three dimensional data structure comprising *rotations per minute*, *orders*, and *noise* in the form of residual values. Regarding the **task abstraction**, we discovered the six tasks (1) *Gain overview of engines*, (2) *Drill-down to engines*, (3) *Identify engine of interest*, (4) *Analyze single engine*, (5) *Assing label to engine*, and (6) *Annotate acoustic measurements*. To qualitatively evaluate IRVINE, we performed a user study with four engineering experts from the domain of acoustic testing for electrical engines. In the context, we uploaded sensor data and gave the users the task *"please find error-prone engines and provide labels and annotations for each engine."* Each study participant was asked to perform the task following the think-aloud method [110], followed by open-ended questions to evaluate how every single view supported the task. To quantitatively assess the usability of IRVINE, we applied the system usability scale [68]. In addition to that, we measured how much labeling speed improved by using IRVINE compared to manual analyses of acoustic signatures of electrical engines.

Results and Implications for Future Studies: Our results suggest a high usability and usefulness of IRVINE as part of the improvement of a real-world manufacturing process. The score from the usability scale was 85, which is well above the average score of 68 [68]. Specifically, with IRVINE domain experts were able to label and annotate produced electrical engines more than 30% faster. In this regard, we not only contributed novel interaction techniques to interactively cluster and label complex data structures but also showed that labels and annotations can be a way to formalize and share expert knowledge. Here, the *Hypermatrix*, as well as specifically designed two colored glyphs, where a new interaction technique to significantly improve labeling as well as knowledge sharing. As an implication for future design studies, engineering experts reported that it is important to analyze data across the manufacturing process connecting several stations. This challenge formed the basis for our next design study project ManEx [33].

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Conceptualized and implemented the system's back-end to automatically transform spectrograms of electrical engines into hypermatrices as well as residual spectrograms.
- Implemented and evaluated a neural network in the form of a self-organizing map to cluster hypermatrices.
- Abstracted the problem domain, the data, and analysis tasks.

- Conceptualized all interaction techniques, for example, the representation of a hypermatrix as a small multiple or the glyph representation of anomalies for electrical engines.
- The conceptualization of novel techniques to interactively label and annotate electrical engines.
- Conceptualized and implemented automatic retraining of the system's underlying self-organizing map to perform additional clustering.
- Designed and developed all interfaces.
- Evaluated the system with four engineering experts.
- Performed an additional evaluation to measure increased labeling speed.

Written Contents Contributed by Myself: Besides parts of the introduction, I wrote the entire contents of the paper, including all tables and figures.

4.1.4 Publication 4: Manufacturing Explorer

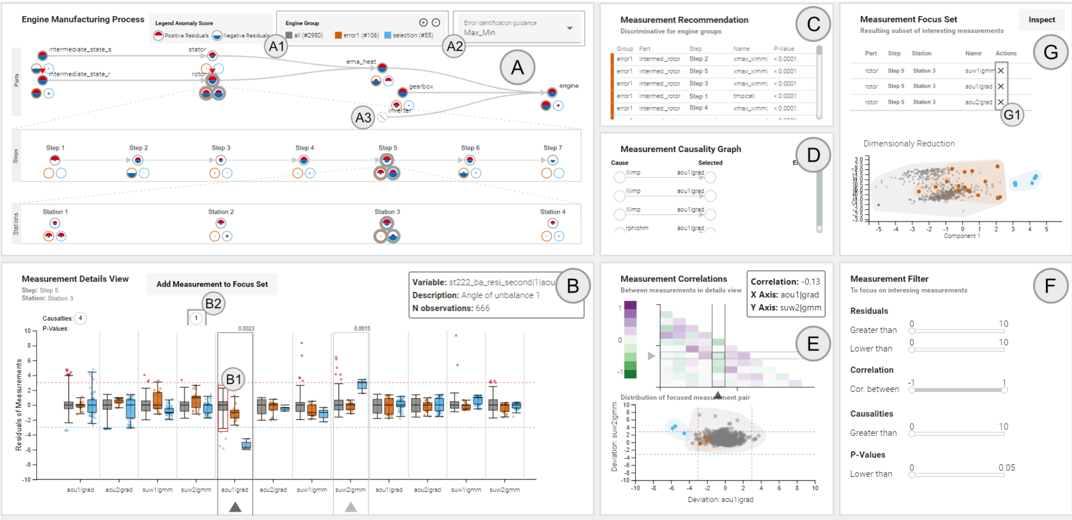


FIGURE 4.4: Screenshot of ManEx. Users have an overview of parts, steps, and stations in (A). Measurement details for selected stations are shown in (B). Users can locate relevant measurements based on t-tests in (C) or causalities in (D). The correlation for a selected sample of measurements is shown in (E). A sample of measurements can be filtered in (F) and the strength of the computed relationships be evaluated in (G).

Citation: *J. Eirich, G. Koutroulis, B. Mutlu, D. Jäckle, R. Kern, T. Schreck, and J. Bernard. Manex: The visual analysis of measurements for the assessment of errors in electrical engines. IEEE Computer Graphics and Applications, pp. 1–13, 03 2022*

Content: While previous design studies focused on data from single stations of the manufacturing process, the goal of this publication was to build a system, which enables the analysis of sensor data across the manufacturing process. To analyze parts across the manufacturing process, engineers are currently hindered by the following three challenges: (1) **Hierarchical challenge:** Engines comprise a complex hierarchical structure of subcomponents. (2) **Temporal challenge:** Locating error causes along the manufacturing processes is a difficult procedure. (3) **Measurement challenge:** Large numbers of heterogeneous measurements impair the ability to explain errors in engines. We address these three challenges with the Visual Analytics system ManEx (Manufacturing Explorer). ManEx provides interactive interfaces to analyze measurements of engines across the manufacturing process. Engineers can use their domain knowledge to analyze anomalies in measurements and judge whether an anomaly serves to identify engines, which might either include errors. In this project, we were inspired by the system IRVINE [28] particularly to show anomalies in the manufacturing process as two colored glyphs. Furthermore, the causal discovery provides insights to find the root cause of detected errors. For a detailed analysis, measurements can be added to a subset of measurements, which we refer to as a focus-set. Engineers can evaluate the quality of a focus-set until it is capable of detecting errors in the manufacturing process.

Research Method: ManEx was designed and developed in the form of a design study project [105], where we also used the nested model by Munzner [84]. A formative evaluation was carried out with three engineering experts from BMW. First, we abstracted the manufacturing process along the **hierarchical** and **temporal** dimension. These two dimensions provided the foundations to abstract manufacturing processes, as shown in Figure 4.5. In terms of the **data characterization**, we abstracted electrical engines into the two categories *engines with explainable error class* and *engines with novel unexplainable error class*. While the former describes engines with an error, that engineers are already familiar with, the latter are engines, which contain an error, that engineers experience for the first time. Furthermore, we distinguish between *known measurements* and *unknown measurements*. The former are measurements, which engineering experts know very well, while the latter are measurements, that engineers do not know before analysis begins (e.g., a measurement from a station the engineers are not responsible for). Considering the task abstractions, we discovered the following six new tasks, which support the analysis of engines and measurements across the manufacturing process: (1) *Lookup engine group and measurement*, (2) *Locate measurement of interest*, (3) *Browse for engine group*, (4) *Explore engines and measurements*, (5) *Compare engine groups*, (6) *Identify relations between measurements*. To qualitatively assess ManEx, we performed a user study with five engineering experts with knowledge about different components of electrical engines. For the study, we uploaded sensor data and gave the users the task *"please find a set of measurements that you believe to be relevant for developing a testing procedure for the specific error"*. Each study participant was asked to perform the task following the think-aloud method [110], followed by open-ended questions to evaluate how every single view supported the task. To quantitatively assess the usability of ManEx, we applied the system usability scale [68]. In addition to that, we evaluated how new focus sets discovered by the engineers served to improve testing procedures in the manufacturing process.

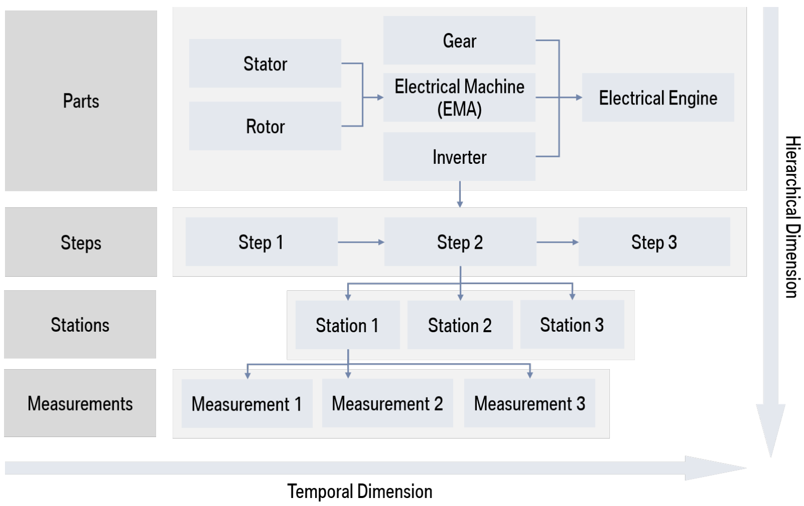


FIGURE 4.5: Assembly of an electrical engine along the hierarchical and temporal dimension.

Results and Implications for Future Studies: Our results suggest a high usability and usefulness of ManEx as part of the improvement of a real-world manufacturing process. The score from the usability scale was 83.5, which is well above the average score of 68 [68]. In this design study, we designed and developed novel interaction techniques to analyze and compare data from groups of produced parts along the manufacturing process. First, we provide a novel two-colored glyph view that facilitates the comparison of produced engines. Next, we map causal graphs directly on the manufacturing process in a *details on demand* [106] manner. Furthermore, we allow interactive evaluation mechanisms of individually created focus sets by means of dimensionality reduction. As an implication for future design studies, engineers reported that it was hard to interpret variables from distinct manufacturing stations without knowing the context. This challenge formed the basis for our next design study project ManKnowVis [31].

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Conceptualized and implemented the system's back-end to automatically connect and process sensor measurements for the entire manufacturing process of electrical engines.
- Conceptualization of novel algorithms for anomaly detection in sensor measurements with residual values.
- Abstracted the problem domain, the data, and analysis tasks.
- Conceptualized a high-level representation of the industrial manufacturing process along a temporal and hierarchical dimension.
- Conceptualized all interaction techniques, for example, the comparison of groups of engines with error to error-free engines with dedicated glyph representations or the projection of causality graphs on top of manufacturing process representations.
- Designed and developed all interfaces.
- Evaluated the system together with five engineering experts.

Written Contents Contributed by Myself: Besides the Section "Causality Computation" I wrote the entire contents of the paper, including all tables and figures.

4.1.5 Publication 5: Manufacturing Knowledge Visualization

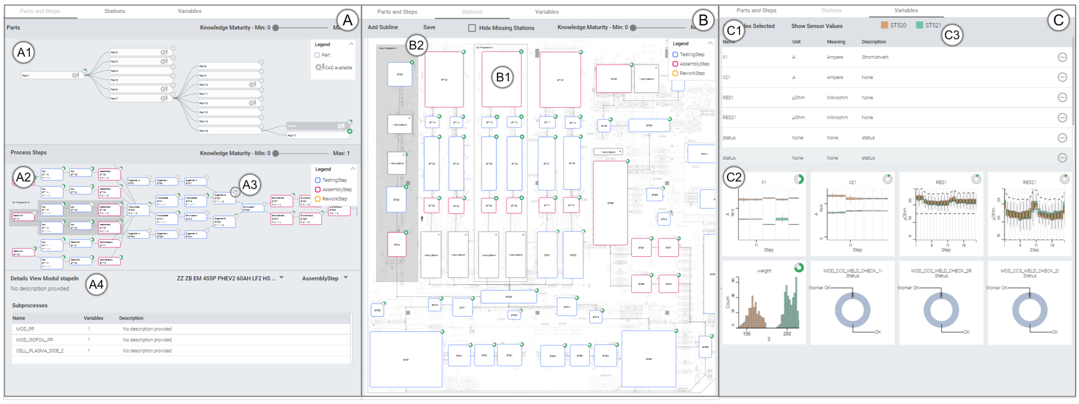


FIGURE 4.6: Screenshot of ManKnowVis, which is divided into the three-level of detail views, high, middle, and low. The *parts and steps* view (A) shows the coarsest, the *station* view (B) the middle, and the *variable* view (C) the highest level of detail. On the user's desktop screen two views are shown side by side, where users can freely switch between each view using the navigation panel on top. (A) shows parts (A1) and steps (A2) and details for both entities in (A4). Knowledge maturity levels are shown as donut charts (A3) in the views A, B, and C. Stations are displayed in (B) and arranged following the physical layout of an assembly line (B1), which is shown as a background image. Sub-Lines are displayed in (B2). To save display space, details for stations are also displayed in (A4). Selected variables for a part, step, or station are displayed in C. Metadata variable attributes are shown in (C1) and their distribution in (C2). Depending on their type, variables are displayed as histograms, multiple boxplots, or donut charts.

Citation: *J. Eirich, D. Jäckle, M. Sedlmair, C. Wehner, U. Schmid, J. Bernard, T. Schreck, and J. Bernard. Manknowvis: How to support different user groups in contextualizing and leveraging knowledge repositories. Preprint*

Content: In previous design study projects [30, 35, 28, 33], we observed a discrepancy between the two stakeholder groups *providers* and *consumers*. Knowledge *providers*, such as engineers, have domain knowledge about specific parts of the manufacturing process. For instance, they know exactly what data is recorded at manufacturing stations. However, *providers* do not have the necessary skills to perform sophisticated data-driven analyses, such as the training of machine learning models. Such analyses are usually performed by highly skilled data experts, such as data scientists. However, this group does not have first-hand domain knowledge about the manufacturing process. Therefore, they rely on consuming knowledge, which is why we refer to them as *consumers*. Instead of enabling *providers* directly via visualization interfaces, the main challenge in manufacturing setups is to bridge the knowledge gap between these two groups and foster fruitful, effective, and seamless collaborations. Hence. as the last design study of this thesis investigated how Visual Analytics can support bridging the gap between different user groups with distinct levels of knowledge. The overall goal was to investigate how Visual Analytics can enable the efficient and easy sharing of already formalized knowledge among distinct stakeholder groups. To do so, we designed and developed the system ManKnowVis (Manufacturing Knowledge Visualization), in which we contextualize data from multiple knowledge repositories of a manufacturing process for battery modules used in electric vehicles (Please note that this project is also part of the publicly funded project KiProQua). We base our study on existing documentation from different data sources, which provide a holistic depiction of an assembly line for battery modules from one of BMW's manufacturing facilities. Furthermore, we were provided with an existing ontology that supported us in contextualizing entities of the analyzed assembly line. *Providers* can use ManKnowVis to enrich existing manufacturing entities with their knowledge by describing and connecting entities. ManKnowVis provides them with visual interfaces to externalize their knowledge, such as grouping stations via annotation functionalities directly on shop floor layout images. *Consumers* can use ManKnowVis to get a detailed understanding of distinct data-driven analyses they are addressing. Therefore, ManKnowVis bridges the gap between *providers* and *consumers* by supporting *providers* in externalizing their knowledge in a systematic and structured way and by providing *consumers* with efficient access to contextualized knowledge, which was previously inaccessible. For *consumers* this knowledge is particularly helpful in understanding complex manufacturing data, which is the necessary foundation to perform data-driven analyses, such as root cause error analyses or anomaly detection to name a few. As a result, ManKnowVis directly influences the success of data-driven analyses.

Research Method: ManKnowVis was designed and developed in the form of a design study project [105], where we also used the nested model by Munzner [84]. First, we performed interviews with seven *consumers* and seven *providers* to to better understand how *consumers* perform data-driven analyses and how *providers* externalize their knowledge. To describe production processes and involved stakeholders, we relied on an existing ontology from BMW. The ontology guided us in developing ManKnowVis and to represent its underlying manufacturing data. A detailed description of how we represent distinct user groups with different knowledge levels together with the ontology is presented in Figure 4.7. Furthermore, we performed a formative evaluation, which guided the system design in close collaboration with two *providers* and *consumers*. Last, we evaluated the final system in the form of a case study [136] with seven participants (three *providers* and four *consumers*). Here, we performed kick-off meetings with all study participants to introduce all features of ManKnowVis. After that participants used ManKnowVis for two weeks. In this regard, we asked *providers* to document entities in the form of describing and connecting entities. *Consumers* were encouraged to analyze entities and evaluate whether they were able to use available information for data-driven analyses. After the study, we interviewed *providers* and *consumers* to evaluate how ManKnowVis supported them with their tasks.

Results and Implications for Future Studies: In this design study, we designed novel interaction techniques to efficiently analyze different knowledge repositories regarding manufacturing processes. In this regard, we compared several visualization approaches to visualize manufacturing entities, which we showed in Figure 4.7. First, we experimented with forced layout graphs, which we then changed to a directed graph. Finally, we further enhanced our visualization design and implemented distinct linked views each representing different ontology entities. Furthermore, we contributed to the theory of design study methodology [105] by taking a multi-user instead of a single user perspective.

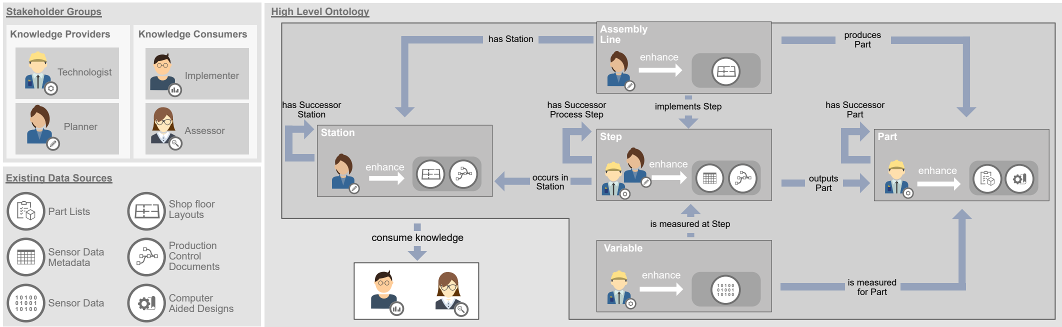


FIGURE 4.7: Relevant stakeholders, data sources, and high-level ontology for ManKnowVis. Technologists and planners provide and implementers and assessors consume knowledge. We use six different data sources as input data for ManKnowVis. A manufacturing process can be described with the entities "Assembly Line", "Station", "Step", "Part", and "Variable". Each entity can be represented with existing data and enhanced by providers. Consumers can inspect this high-quality process knowledge to improve analysis outcomes.

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Conceptualized and implemented the system's back-end to automatically map different entities of a knowledge graph with each other.
- Conceptualized different user roles with distinct levels of knowledge regarding industrial manufacturing processes into an integrated framework.
- Abstracted the problem domain, the data, and analysis tasks.
- Extended the design study methodology to account for a multi-user perspective.
- Conceptualized all interaction techniques, for example, the mapping of sensor measurements to part computer-aided designs with annotations or the connection of distinct knowledge graph interfaces with individual interfaces.
- Designed and developed all interfaces.
- Evaluated the system in the form of a case study with seven engineering and data experts.

Written Contents Contributed by Myself: Besides parts of the introduction, I wrote the entire contents of the paper, including all tables and figures.

4.2 Theoretical Frameworks

Apart from our five design study projects, we also contributed theoretical models, which we describe in the following subsections. Each subsection is structured as follows. First, we provide an overview of the content of each individual model. Next, we provide insights into each research approach for the publications. This is followed by a brief summary of our results and implications for research as already outlined in Figure 3.4. Last, I report on scientific contributions and written content in each publication of myself. For details of each publication please refer to Chapter 6.

4.2.1 Publication 6: Organizational Knowledge Creation Visual Analytics

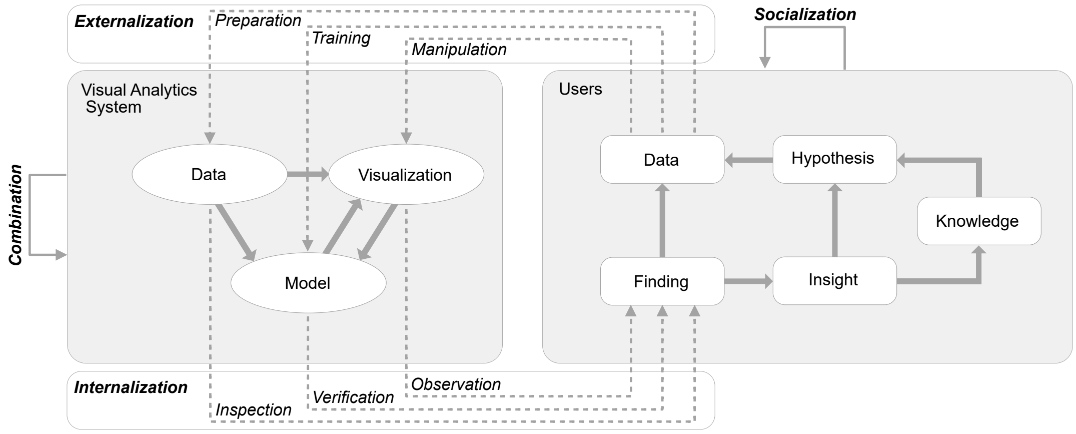


FIGURE 4.8: Conversion between tacit and explicit knowledge in VA. Gray boxes indicate components from Sacha et al. [100], where we also use the SECI phases from Nonaka and Takeuchi [88].

The publication got the award for best track paper on the track for Knowledge Management and Digitization at the European Conference on Information Systems (ECIS).

Citation: *J. Eirich, D. Jäckle, S. Werrlich, and T. Schreck. Visual analytics in organizational knowledge creation: A case study. In European Conference on Information Systems, 04 2021*

Content: This study was above all motivated by our first Visual Analytics system VIMA [30]. During the development of VIMA, we experienced that there was an available theoretical model to describe Visual Analytics systems in the context of organizational knowledge creation. As a result, with this publication, we extended the work of others who have explored aspects of organizational knowledge creation in the context of Visual Analytics. To do so, we merged the model of knowledge creation in Visual Analytics from Sacha et al. [100] with the model of organizational knowledge creation from Nonaka and Takeuchi [88] into an integrated theoretical model. Our goal was to understand and conceptualize the process of knowledge creation and transformation in the presence of a complex Visual Analytics system that proactively augments decision-making and cognitive reasoning. The theoretical framework is shown in Figure 4.8.

Research Method: To develop our theoretical framework, we performed a literature review about knowledge creation in Visual Analytics and organizational knowledge creation. To evaluate the resulting framework, we carried out a case study [136] over a period of nine months. We collected machine sensor data of the manufacturing process, that underlie VIMA, face-to-face semi-structured interviews, and direct observations. The sensor data were used as input for VIMA, to enable experts to conduct multiple explorative analyses. Additionally, we conducted seven semi-structured interviews with four domain experts for electrical engines and three with electrical energy storage systems experts. Furthermore, an on-site researcher-made direct observations on the studied processes, and recorded his findings following the think-aloud method [110] on a weekly basis.

Results: From the case study we derived new mechanisms that have to be considered for organizational knowledge creation, which we describe in the following. During the interaction between individuals, information is interpreted by each individual to become knowledge. VIMA itself is not able to create knowledge alone. It rather transforms available data into information and interprets it in a way to create meaning. Through the interaction of domain experts with the system and each other, we observed a transformation between tacit and explicit knowledge that enhanced the knowledge base of the users and improved the system. Nonaka and Takeuchi [88] argue that socialization is a result of direct social interaction, for example, individuals, who exchange tacit knowledge through hands-on learning experiences. In our case, socialization was triggered through system interactions, causing ripple effects of knowledge creation for both individuals using the system and those not using it. Since after multiple analyses with the system, users approached other colleagues to verify assumptions or discuss findings, VIMA mediated the exchange of tacit knowledge between individuals. Traditionally, tacit knowledge is externalized consciously through physical interaction, articulated through images, symbols, and language. In turn, VIMA facilitates the externalization of tacit knowledge, e.g. by requiring users to label data. It then combines various labels into larger collections of externalized knowledge, which are then transformed into recommendations and visualizations that are supplied to a wide audience of users. This combination of continuous externalization, automated combination, and visualization affects the internalization of knowledge by VIMA's users. Considering the process of combining multiple different sets of explicit knowledge is recombined and reconfigured manually and often contains comparable small datasets. In turn, we observed that combined efforts were mostly carried out by VIMA, where huge datasets were recombined into smaller better comprehensible ones, that were easily interpretable by its users. Nonaka and Takeuchi [88] describe the process of internalizing explicit knowledge as similar to "learning by doing". Here, an individual is driven by the inherent wish to learn something new. Instead, while experts conducted observations, verifications, and inspections of the data, we observed a learning process where explicit knowledge from data, models, and the visualizations was transformed into new tacit knowledge, such as changed mental models considering user's problem domains.

As a result, we developed the following four propositions that can guide future research in investigating similar areas of knowledge creation and Visual Analytics.

- *New tact it knowledge is created through socialization between individuals as a result of system interactions with one user and a VA system, from findings, insights, actions, or hypotheses.*
- *Tacit knowledge is converted into explicit knowledge by means of externalization through VA system interactions via manipulating visualizations, training models, or preparing data, and are articulated through features, labels, and models.*
- *New explicit knowledge is created through the combined abilities of VA systems, in databases, models, or visualizations.*
- *Explicit knowledge is converted into tacit knowledge in internalization as a result of inspecting data, verifying models, or observing visualizations in VA systems.*

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- The analysis of literature for organizational knowledge creation and knowledge creation in visual analytics.
- The development of a theoretical model to conceptualize and describe mechanisms of organizational knowledge creation with visual analytics.
- The discovery of novel mechanisms for organizational knowledge creation afforded through visual analytics.
- The evaluation of the model in the form of a case study with seven engineering experts.

Written Contents Contributed by Myself: I wrote the entire contents of the paper, including all tables and figures.

4.2.2 Publication 7: Labels in organizational learning

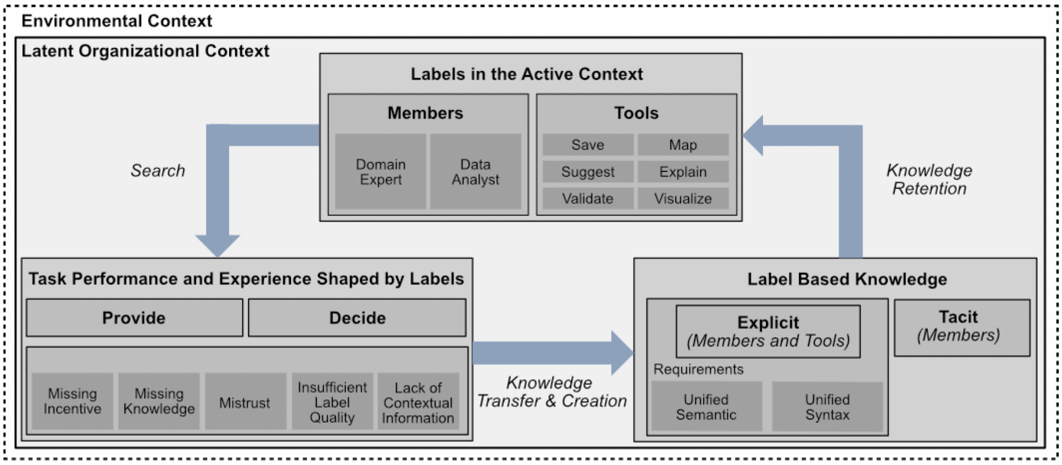


FIGURE 4.9: The organizational learning framework from Argote and Miron-Spektor [7] contextualized to labels.

Citation: *J. Eirich and F.-P. Diana. The life cycle of data labels in organizational learning: A case study of the automotive industry. In European Conference on Information Systems, 04 2022*

Content: Based on the findings from the Visual Analytics systems RfX [35], IRVINE [28], and ManEx [33], this study was motivated to better understand how labeled data flows through organizations and how a label can be related to the creation, formalization, and transfer of knowledge. In general, labels describe the class of a data instance (e.g., an image containing a *cat* or a *dog*). However, in publications 6 [32], we also found out that labels represent a form of valuable and already externalized expert knowledge (e.g., an electrical engine that already was analyzed by an expert and classified as *not ok*). Hence, in this study, we draw on the theoretical framework of organizational learning by Argote and Miron-Spektor [7] and contextualize it to the use of labels in organizations. In particular, we analyze how labels that are used in the three Visual Analytics systems RfX [35], IRVINE [28], and ManEx [33], affect organizational members' tasks and knowledge retention. We choose these three Visual Analytics systems because they all rely one way or another on labeled data. RfX [35], needs labels for the training of its underlying random forest. IRVINE [28] is a Visual Analytic system specifically designed to create labeled data. One core component of ManEx [33] is to compare measurements of groups of labeled parts. Our results contribute a contextualized theoretical framework, which helps to understand the role of labels in organizations. Furthermore, we provide propositions on how labels affect organizations and outline possible future research opportunities in this context.

Research Method: As a research method, we applied a case study [136], which we conducted over sixteen months. During this period, we collected and analyzed data from three sources. (1) Semi-structured interviews: We interviewed 15 participants—seven data analysts and eight engineering experts from BMW. (2) Direct observations: Furthermore, we investigated how organizational members use the three Visual Analytics systems (RfX [35], IRVINE [28], ManEx [33]) following Someren et al.'s [110] think-aloud method. The observations helped us to understand how organizational members (e.g., domain experts from different organizational units) share and create knowledge in the form of labels. Furthermore, the observations helped to contextualize label-based tasks. (3) Documentation from design and development iterations: Each of the three Visual Analytics systems, was designed and developed in several iterations. During each iteration, we documented numerous types of information. This information included the tasks users perform while using the tools or Visual Analytics requirements used as design targets for each tool. This data helped to abstract high-level system requirements, which Visual Analytics systems need to address in the context of labels. The data also informed the definition of general tasks, which members and tools perform when using label-based systems.

Results: As a result of our case study, we contextualized the the original framework for organizational learning from Argote and Miron-Spektor [7] with labels. The resulting framework is shown in Figure 4.9. We structured our results along the three dimensions, *active context*, *task performance and experience*, and *knowledge*. Regarding the *active context*, we found out that two groups of organizational members heavily interact with each other, which are data analysts and domain experts. Furthermore, we derived the six high-level requirements *save, suggest, validate, map, explain,* and *visualize* that Visual Analytics systems should address when being embedded into the context of labels. Regarding *task performance and experience*, we discovered two major tasks that members perform when working with labels. The two tasks are to *provide* labels or to *decide* based on labels. Furthermore, we observed that these tasks are currently hindered by the five impediments *missing incentive, missing knowledge, mistrust, insufficient label quality*, or a *lack of contextual information*. In terms of the dimension *knowledge*, we found out that labels impact the definitions of *explicit* and *tact* knowledge. First, regarding *explicit* knowledge, labels represent a valuable formalized product of knowledge, which can be shared in organizations. To be shared efficiently, however, labels need a unified semantic and syntax. In addition to that, we found out that organizational members increased their *tacit* knowledge when working with labels. To evaluate our findings in future work, we proposed seven research questions, which can be addressed by researchers, who investigate similar label-related domains. The research questions are the following:

- RQ1: How can labels efficiently be used across different organizational units and between organizations?
- RQ2: How must label-based tools be designed to improve knowledge creation and transfer across different organizational units and between organizations?
- RQ3: How can organizations design incentives to foster label-based tasks and roles?
- RQ4: How can organizations overcome impediments related to labels that hinder effective task performance?
- RQ5: Is a data labeler a new position that organizations need to incorporate? And if so, what are the tasks, and what knowledge is required of a labeler?
- RQ6: How must labels be modeled in organizational ontologies to support organizational knowledge management?
- RQ7: How can ontologies support the development of label-based tools?

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- The analysis of literature in the field of labeling and organizational learning.
- The development of a theoretical model to describe and conceptualize the role of labels in organizational learning.
- The discovery of novel label-related tasks in organizations, knowledge creation, formalization, and sharing mechanisms in the form of labels, and label-related users and tools in organizations.
- The development of propositions and research questions to support future research in the context of labels.
- A case study with 15 engineering and data experts to develop the theoretical model.

Written Contents Contributed by Myself: The paper was written in close collaboration with the second author, where I wrote the main parts of the paper and hence became the first author.

4.2.3 Publication 8: Visualization Design Principles for Manufacturing Processes

	Aim	Mechanism	Rationale
Data component	***DP-D-1.1 Principle of Data Contextualization***		
	Contextualize data from all distinct data sources	Use a common semantic to contextualize data	Providing a contextualized data set results in a big picture of the quality of produced parts
	DP-D-1.2 Principle of Data Filtering		
	Reduce the amount of data that must be analyzed	Include interactive filtering mechanisms	Removing unimportant information reduces analysis time
	DP-D-1.3 Principle of Parts of Special Interest		
	Compare groups of parts with an error to parts with no error	Enable the upload of parts of interest	Comparing two parts groups allows the early detection of faults
Model component	***DP-M-1.1 Principle of Data Harmonization***		
	Compare measurements from different sources	Compute residual values from the measurements	Computing residual values make measurements comparable
	DP-M-1.2 Principle of Hypothesis Testing		
	Identify measurements of interest for specific parts	Compute two-sided t-tests for groups of parts	Computing t-tests helps to identify interesting measurements
	DP-M-1.3 Principle of Focus-Set Selection and Validation		
	Build and validate a subset of measurements	Apply dimensionality reduction to the subset	More than three dimensions of measurements cannot be visually represented
Visualization component	***DP-V-1.1 Principle of the Hierarchical Representation***		
	Visualize hierarchy of parts, steps, and stations	Use a tree representation	Tree views allow identifying errors for parts, steps, and stations
	DP-V-1.2 Principle of the Temporal Representation		
	Visualize temporal dimension of part assembly	Use a directed graph	Directed graphs situate relations of parts, steps, and stations
	DP-V-1.3 Principle of Anomaly Abstraction		
	Enable the identification and assessment of anomalies	Provide different anomaly metrics and views	Different anomaly abstractions and metrics allow assessing anomalies and their causes

FIGURE 4.10: Design principles for the design and development for Visual Analytics system in industrial manufacturing settings.

Citation: *J. Eirich. Visual analytics for IoT data from large-scale manufacturing processes. American Conference on Information Systems, 08 2022*

Content: This study was motivated by the design study project for the Visual Analytics system ManEx [33]. One contribution of this design study was the contextualization of manufacturing processes along the hierarchical and temporal dimensions (See Figure 4.5. Based on this abstraction, the goal of this study was to develop general design principles to design and develop Visual Analytics systems in industrial manufacturing settings. To do so, we reduced the functionality of ManEx to its core features, excluding for example the causality graph computation.

Research Method: We performed this study following a design science research approach [50, 117] where we performed the study along the five phases (1) *Awareness of the Problem*, (2) *Suggestion*, (3) *Development*, (4) *Evaluation*, and (5) *Conclusion*.

For phase (1), we drew on existing literature on IoT data in the manufacturing industry, Visual Analytics in general, and Visual Analytics applications in the manufacturing industry. Furthermore, we carried out semi-structured interviews with six engineering experts working at BMW. In phase (2), we used the results from the literature analysis, the findings from our design study about ManEx [33], implications from the other four Visual Analytics system VIMA [30], RfX [35], IRVINE [28], ManKnowVis [31], and the interviews to develop nine design principles, which represent normative design decisions that provide guidance on how to design Visual Analytics systems in manufacturing settings. In phase (3), we used the design principles to redesign ManEx to its core features. In phase (4), we performed both a qualitative and quantitative evaluation of ManEx relying on the think-aloud method [110] in combination with the system usability scale [68]. The qualitative evaluation was carried out with five engineering experts. In phase (5), we concluded the project with the generalization and conceptualization of our results in publication 8. Our results suggest a high usability and usefulness of the adapted ManEx system.

Results: The score from the usability scale was 83.5, which is well above the average score of 68 [68]. Furthermore, the design principles are shown in Figure 4.10. Each design principle is structured along with the three components of Visual Analytics, which are the data, model, and visualization component according to Keim et al. [57] (See Figure 2.3). These design principles can inform future researchers in designing similar Visual Analytics systems in manufacturing contexts. For example, we found out that the principle of data harmonization in the model component of Visual Analytics systems is important to be able to compare measurements from different sources (e.g., voltages against temperatures). This can be addressed by computing residual values from measurements of produced parts, as we did in the design studies VIMA [30], or IRVINE [28].

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- The conceptualization of design principles to guide Visual Analytics system design to analyze IoT data from industrial manufacturing processes along the model of Keim et al. [57].
- The evaluation of the design principles with five engineering experts.

Written Contents Contributed by Myself: I wrote the entire contents of the paper, including all tables and figures.

4.2.4 Publication 9: Identification of Anomalies in Highly-Integrated Electric Drives

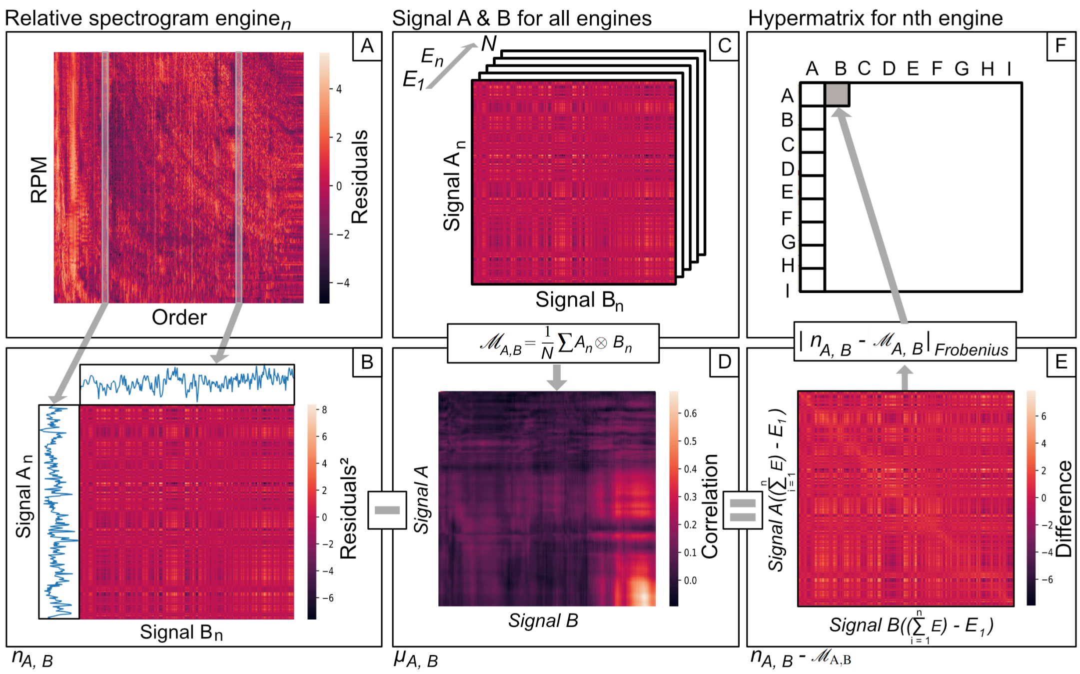


FIGURE 4.11: Computation of our Hypermatrix. From a given spectrogram in (A), we extract two columns that are correlated with each other in (B). This is done for each signal pair over all engines in (C). Next, we subtract the resulting mean signal combination in (D) from each signal combination of a single engine which results in the deviation of a signal pair from one engine to all other engines in (E). Each signal pair is then aggregated and stored in a new matrix in (F).

Citation: *J. Bonart, J. Eirich, P. Hümmer, and W.-G. Drossel. Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms. In IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics, and Drives, 08 2021*

Content: This work was the initial basis for our design study project about the Visual Analytics System IRVINE [28]. The goal of this publication was to develop a new metric for identifying anomalies in electrical engines by evaluating their vibration signatures. To do so, we developed a method that scans engine data in the form of spectrograms, transforms them into reduced difference-correlation matrices, and thus takes secondary error mechanisms into account when analyzing anomalous engines. In the design study project IRVINE [28], we refer to this newly developed abstraction as *´´hypermatrix"*. The methods extend the state-of-the-art anomaly detection metrics in the field of acoustic testing, such as euclidean metrics or residual metrics, which only take primary error symptoms into account.

Research Method: The study was performed in an experiment setup, where we analyzed spectrograms from a sample of 600 electrical engines measured on a test bench. First, we evaluated how hypermatrices serve to analyze anomalies compared to established metrics, such as the euclidean metric and the residual metric. As a next step, we evaluated whether hypermatrices are able to detect the two fault classes (1) shorted windings in electrically excited rotors and (2) faulty c-bearing of an electrical engine.

Results: First, we found that hypermatrices outperformed the euclidean and residual metrics, which are well-established metrics to analyze spectrogram data. Second, we showed in a real-world use case how hypermatrices serve to better and more efficiently detect errors in electrical engines by considering primary and secondary excitation mechanisms. Due to the success of the computation of hypermatrices, we took this metric as starting point to develop the Visual Analytics system IRVINE [28] to facilitate the analysis of a large number of electrical engines by analyzing their acoustic signatures.

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Contributed to the development of the hypermatrix.
- Evaluated how this metric could be leveraged for the future Visual Analytics system IRVINE [28].

Written Contents Contributed by Myself: I contributed to the conceptualization of the publication's underlying methodology to abstract acoustic signatures in the form of a hypermatrix, designed Figure 3 of the publication, and reviewed the publication.

4.2.5 Publication 10/11: Use Cases of Artificial Intelligence and Visual Analytics in Industrial Manufacturing

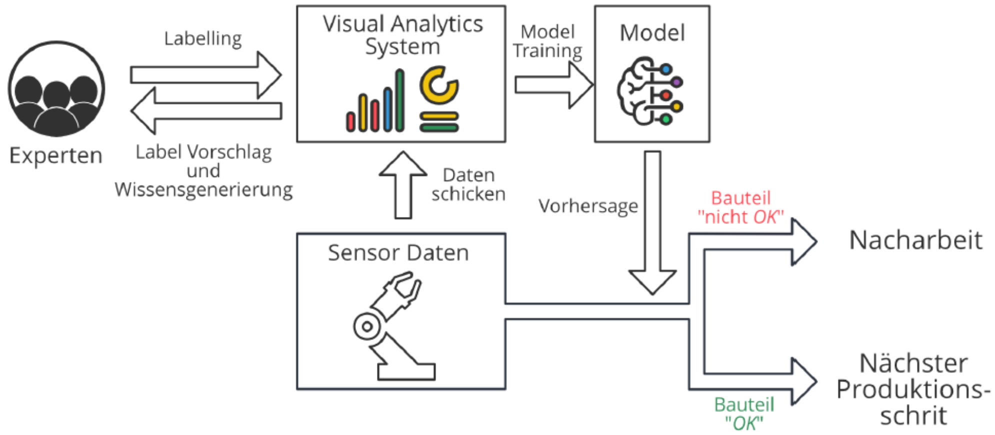


FIGURE 4.12: High level process how Visual Analytics systems can support domain experts in generating, formalizing, and sharing knowledge.

Citations:

- 1) A. Luckow, **J. Eirich**, and Demtroeder. K. Kuenstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie. In *Wie funktionier Data Science?*, pp. 1–11. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.
- 2) **J. Eirich**, A. Luckow, and Demtroeder. K. Visual Analytics zur Optimierung von Pruefprozessen in der Serienfertigung elektrischer Motoren. In *Wie funktionier Data Science?*, pp. 1–7. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

Content: Both publications started as single book chapter. However, the editor requested to separate them into two book chapters. Nevertheless, we will summarize both chapters in this sub-section. The first publication provides a summary of machine learning-related use cases in the automotive industry. It also shows challenges, such as missing labels or data, and provides directions on how such challenges can be solved for example by leveraging labeling systems. The second publication gives an overview of how Visual Analytics can help to create, formalize, and transfer knowledge as shown in Figure 4.12.

Research Method: Both publications do not follow a specific research method. They rather summarize the work of our experience over several years regarding applications of machine learning and artificial intelligence as well as Visual Analytics in the automotive industry.

Results: Both publications have the goal to give a big picture about data science and its applications in real-world use cases. Thus, the results from both publications are not new findings in the field of information visualization or machine learning but rather have to goal to make data science more accessible and easy to understand for novel scholars in this research area.

My Scientific Contributions: Within the scope of this dissertation, I made the following scientific contributions, which are published in this paper:

- Summarized the value of Visual Analytics in manufacturing settings.
- Demonstrated successful use cases of Visual Analytics applications in manufacturing settings.

Written Contents Contributed by Myself: I wrote the entire contents of the book chapter ´´Visual Analytics zur Optimierung von Pruefprozessen in der Serienfertigung elektrischer Motoren", including all tables and figures. Regarding the publication "Kuenstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie", I wrote the Chapter 2.3 an reviewd the rest of the paper.

Chapter 5

Summary and Discussion

This thesis tackles the overall research questions *´´How can Visual Analytics support the creation, externalization, and transfer of knowledge in manufacturing processes?"*. Since, this research question is rather broad, in Section 1.1, we divided it into the four sub research questions:

- "**RQ1:** *How can Visual Analytics support domain experts in the analysis of large-scale manufacturing IoT data to create new knowledge?"*
- "**RQ2:** *How can Visual Analytics support externalizing tacit knowledge."*
- "**RQ3:** *How can formalized knowledge from Visual Analytics be used to improve a Visual Analytics Systems to further support expert analyses."*
- "**RQ4:** *How can created knowledge efficiently be transferred with Visual Analytics among distinct stakeholder groups involved in manufacturing processes?"*

In this chapter, we discuss how each of our publications addressed the four research questions. Apart from that, we summarize the scientific and practical contributions from our publications. Next, we give an overview of the limitations and a future research agenda of this thesis and finish with a conclusion.

5.1 Research Questions

Figure 5.1 summarized how all publications of this thesis helped in answering our initial four research questions.

In general, all publications can be categorized into three categories. The publications **P6, P7** and **P8** are all theoretical approaches, **P1, P2, P3, P4** and **P5** Visual Analytics systems, and **P9, P10** and **P11** other approaches. Here, **P9** addresses the field of acoustic testing while **P10** and **P11** are general application scenarios of Visual Analytics systems in industrial manufacturing setting.

Furthermore, Figure 5.1 follows our general outlined research methodology from Figure 3.4 in Section 3.2. In this regard, the arrows indicate how findings from one publication influenced other publications (also already outlined in Chapter 4).

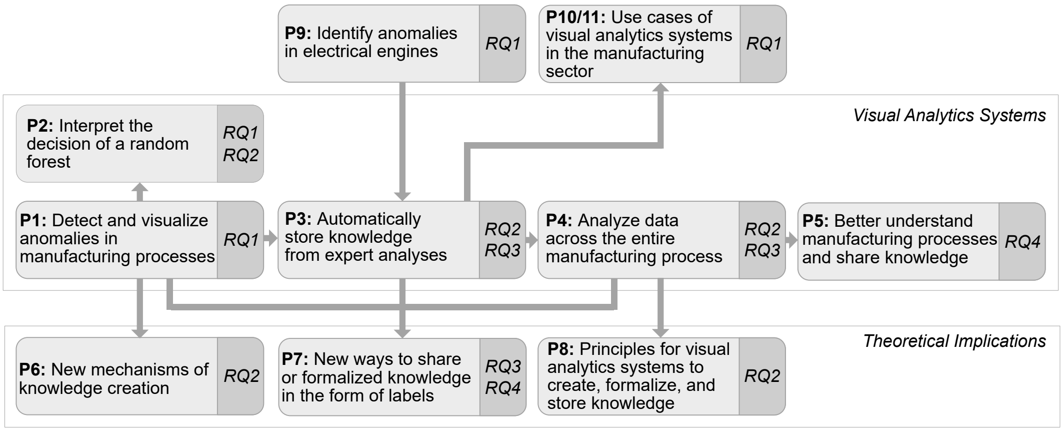


FIGURE 5.1: Publications of this thesis addressing the presented research questions of this thesis.

The first research question about how Visual Analytics can support experts in data analyses of industrial manufacturing data (**RQ1**) is addressed by the two Visual Analytics systems VIMA **P1** and RfX **P2**. VIMA is mainly about detecting and understanding anomalies in serial manufacturing processes while RfX makes thousands of decision trees of a random forest analyzable. One contribution of **P9** is the computation of the hypermatrix - a visual abstraction of high dimensional acoustic sensor data. Even though **P9** addressed the field of acoustic testing, the hypermatrix was one of the motivations to design and develop the Visual Analytics system IRVINE **P3**, which is why we consider this publication also a contribution to **RQ1**. Furthermore, in the two book chapters **P10** and **P11**, we demonstrate how Visual Analytics systems serve to support domain experts in their data analysis tasks in serial manufacturing settings and hence consider these two publications also as a contribution to **RQ1**.

The second research question about how Visual Analytics systems can support the externalization of knowledge is addressed by the three Visual Analytics systems RfX **P2**, IRVINE **P3**, and ManEx **P4**. In RfX, knowledge is externalized via rules derived from decision trees in a random forest. IRVINE is a dedicated labeling system, where expert knowledge is externalized via labels and annotations. In ManEx, domain experts externalize their knowledge by discovering causal relationships along the entire manufacturing process of electrical engines. The externalization of tacit knowledge via Visual Analytics systems is also contextualized in **P6**. In this publication, we discovered explicit knowledge externalization mechanisms afforded through Visual Analytics. For example, one externalization mechanism we discovered is that knowledge is externalized consciously and unconsciously and articulated through features, labels, and models. Furthermore, in **P8** we outline design principles that guide the design and development of Visual Analytics systems for dedicated knowledge externalization tasks. For example, we show that experts can externalize their knowledge with the system ManEx via dedicated measurement sets that help to better detect occurred manufacturing errors.

The third research question about how formalized expert knowledge can be used to improve Visual Analytics systems and thus expert analyses is addressed by the two Visual Analytics systems IRVINE **P3** and ManEx **P4**. Externalized knowledge in IRVINE can be used to retrain its underlying self-organizing map or to make label suggestions in new analyses. In this regard, the more knowledge is provided in IRVINE the better the suggestions for new labels become. In the system, ManEx, externalized knowledge in the form of labeled engines facilitate the detection of hidden relation along the manufacturing process. Again the higher the quality of labels and the more labeled engines are available, the more reliable the results of ManEx regarding the detection of new errors are. Regarding theoretical contributions, in **P7** we show how labels are related to organizational learning and serve to improve Visual Analytics systems. For example, our contextual model gives examples of how a unified semantic and syntax of labels help to improve Visual Analytics systems.

The fourth research question about how Visual Analytics also supports the efficient transfer of created and formalized knowledge is particularly addressed in the Visual Analytics system ManKnowVis **P5**. The purpose of this system is to formalize expert knowledge about the manufacturing process of battery modules. This knowledge is then readily accessible to other stakeholders. For example, domain experts can connect specific manufacturing stations in light of their knowledge. This information is especially relevant for data experts with the goal to perform data analyses along with different manufacturing stations. Furthermore, in **P7** we demonstrate how externalized knowledge in the form of labels can be transferred between data and domain experts.

5.2 Scientific and Practical Contributions

The contributions of this thesis are aligned with the research questions from Section 1.1. In the following sub-sections, we provide a detailed description of how our contributions can inform theory and practice. For details on contributions per paper, the reader is referred to the paper overview from Chapter 4 or the publication itself from Chapter 6.

5.2.1 Theory of Knowledge Creation

So far there exist several theoretical models about knowledge creation or organizational learning. In this thesis we leveraged the well-established models for organizational knowledge creation from Nonaka and Takeuchi [88], the process of organizational learning from Argote and Miron Spektor [6], and the knowledge generation model in Visual Analytics from Sacha et al. [100].

As a first contribution to the theory of knowledge creation, we combined the model of knowledge creation in Visual Analytics [100] with the model of organizational knowledge creation [88] into an integrated theoretical model of organizational knowledge creation with Visual Analytics [32]. Furthermore, we derived new mechanisms for organizational knowledge creation afforded through visualization regarding the creation and conversion of knowledge through socialization, externalization, combination, and internalization. Researchers can build on our findings to evaluate how knowledge in organizations is affected by Visual Analytics systems.

As a second contribution to the theory of knowledge creation, we enhanced the model of organizational learning [6] with the knowledge product labels. In this regard, labels represent an aggregation and conservation of expert knowledge from complex data analysis tasks, such as the evaluation of errors in electrical engines. Hence, labels provide an excellent opportunity to externalize and transfer knowledge in organizations. Researchers can use our enhanced model to evaluate how different stakeholder groups, such as domain experts or data analysts are affected by labels, how labels flow through different organizational departments, or how organizations can leverage labels to retain knowledge.

5.2.2 Visualization Techniques

During the course of this thesis, we experimented with various visualization techniques to support the creation, externalization, and transfer of knowledge with Visual Analytics. A brief summary of the most relevant contributions to information visualization is outlined in Figure 5.2. In (A), we showed that the projection of labeled machine sensor data into a three-dimensional space actually proved to be better than a 2D projection (A1). Furthermore, we showed how statistical outliers can be easily visualized via heatmap and line chart visualizations (A2). In (B), we developed novel methods to show confusion matrices (B1), icicle plots (B2), and node-link diagrams (B3) to better interpret a random forest. In (C), we developed the visual abstraction of a hypermatrix (C1) to better analyze correlation patterns of acoustic signatures. Furthermore, we developed small multiples (C2) as a visual abstraction of hypermatrices. Finally, we provided a novel annotation mechanism (C3) to annotate error causes directly in the raw data. In (D), we developed a novel way to map causality graphs on top of abstraction of manufacturing processes (D1) and provided a new way to compare parts of produced electrical engines with glyphs (D2). In (E), we contributed a novel way to map entities from knowledge graphs on a manufacturing process (E1). Furthermore, we contributed a way to map different entities of knowledge graphs between multiple views in (E2).

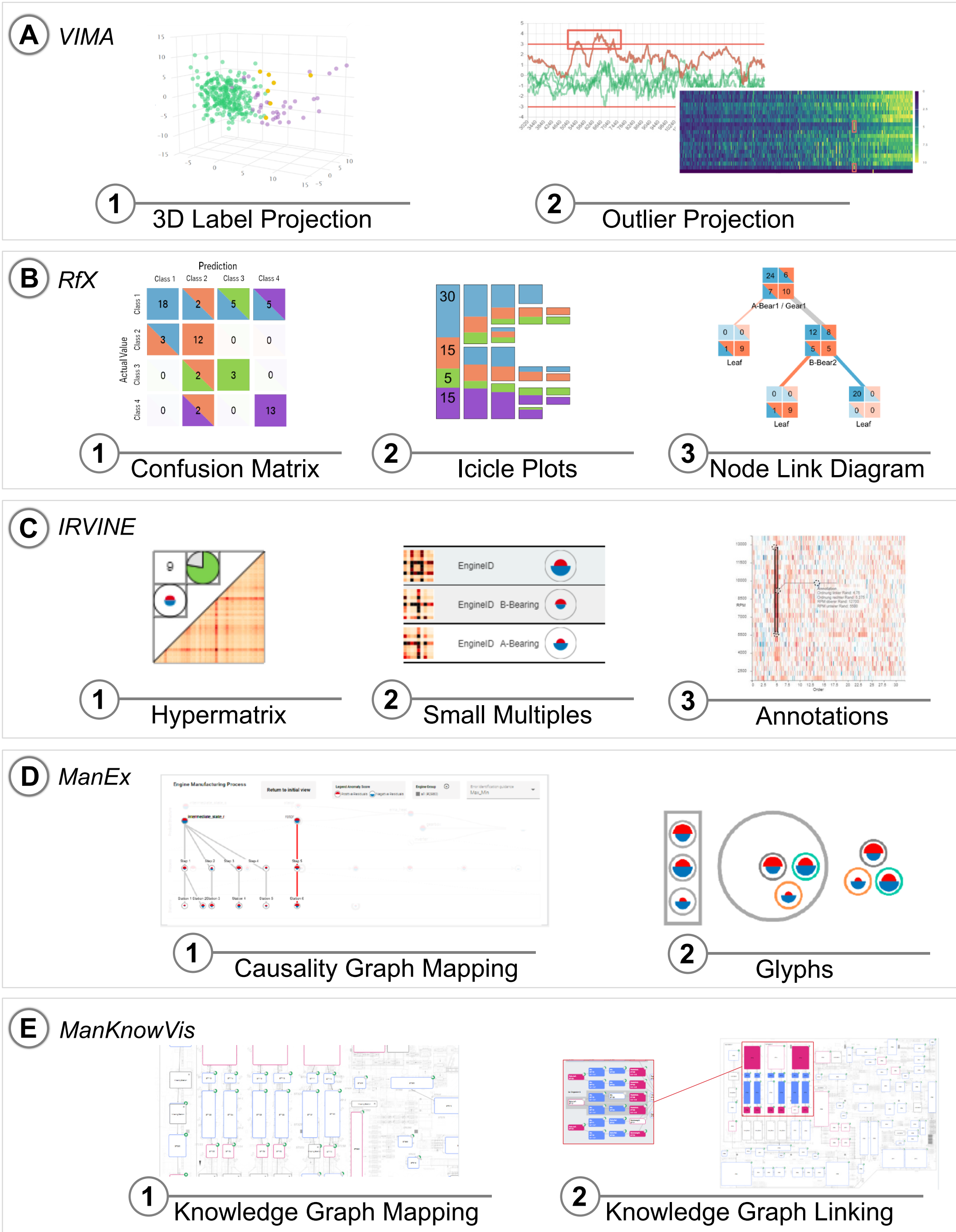


FIGURE 5.2: Visualization contributions from the Visual Analytics systems VIMA, RfX, IRVINE, ManEx, and ManKnowVis.

5.2.3 Abstractions of Manufacturing Processes

We also contributed abstractions of manufacturing processes that researchers can use to build applications in industrial manufacturing settings. For example, we discovered that manufacturing processes can be structured along a hierarchical and temporal dimension. An example of the manufacturing of electrical engines is outlined in Figure 4.5. On the temporal dimension, the assembly of parts is represented along with steps and stations. The hierarchical dimension represents the hierarchy of parts, steps, stations, and measurements. We found that other Visual Analytics systems in the manufacturing sector, such as ViDX [133] particularly target temporal dimensions of the manufacturing process taking only stations and steps into account. However, taking the hierarchical dimension into account also allows us to see the relation of produced parts while considering their assembly. As an additional abstraction of manufacturing processes, we provided an ontology that serves as a semantic layer to describe manufacturing processes, as outlined in Figure 4.7. Researchers can use this ontology as general contextualization of a manufacturing process to design similar Visual Analytics systems as we did in this thesis. The ontology also shows different data sources, which can be relevant to building Visual Analytics systems in manufacturing settings, such as shop floor layouts, or computer-aided designs.

5.2.4 Algorithms to Process Manufacturing Data

Another contribution, which we derived from publications, is algorithms, which serve to process manufacturing data. For example, in the Visual Analytics systems VIMA [30] or ManEx [33], we showed how the computation of residual values serves well to detect outliers from manufacturing data. Furthermore, the computation of hypermatrices [13, 28] also proved to be an excellent metric for a holistic analysis of acoustic sensor data for electrical engines. Finally, in the publication RfX [35], we developed a semantic and syntactic similarity score, which helps to identify representative decision trees from a random forest. Identifying representative decision trees from a random forest surpasses the manufacturing context and can be applied to any type of random forest. Thus, researchers can leverage these two metrics in application scenarios beyond industrial manufacturing processes.

5.2.5 Application and Enhancement of Established Theoretical Models

Another contribution of this thesis is the application and enhancement of established models to real-world scenarios. For example, in the design study project VIMA [30], we applied the model of *visual interactive labeling* [12] as well as the model of *knowledge creation in Visual Analytics* [100]. Another example, is the combination of the *SECI* model [88] with the model of *knowledge creation in Visual Analytics* [100] into an integrated theoretical model to describe knowledge creation in Visual Analytics in broader organizational contexts [32].

5.2.6 Design Study Methodology

Regarding the design study methodology, we showed how this research design fits well to address real-world problems. Apart from that, we also see one contribution of this thesis is an application of design studies to multi-stakeholder scenarios as shown in Figure 4.7. So far, this aspect is not covered so far by Sedlmair et al's. [105] original methodological framework. Thus, we see especially the design study project ManKnowVis [31] as an exploratory step towards this direction. Finally, in all design study projects of this thesis, we contributed several domain, task, and data abstractions about the problem domain of industrial manufacturing processes, which researchers and practitioners can rely on to tackle similar domain problems. For example, we believe that our task abstractions of the labeling system IRVINE [28] can be applied to any labeling scenario in industrial manufacturing scenarios where three-dimensional data structures, such as spectrograms are present.

5.2.7 Design, Development, and Evaluation of Five Novel Visual Analytics Systems

Furthermore, we contributed to the theory and practice with the reporting of the design, development, and evaluation of our five presented Visual Analytics systems. During the design and development of each system, we provided domain, task, and data abstractions of industrial manufacturing settings. Furthermore, we outlined common challenges and pitfalls when designing Visual Analytics systems in manufacturing settings, such as the knowledge discrepancy between engineering and data experts. This documented knowledge in our publication list can help researchers and practitioners to tackle similar domain problems and to avoid the challenges and pitfalls, we outlined in our design studies. For example, we experienced from several data analytics projects in our domain that due to their robustness data experts preferred sophisticated machine learning models, such as neural networks or random forests over simple white-box approaches, such as rule-based recommendations. However, we also observed that engineering experts rather prefer rule-based recommendations over black-box modeling approaches. As a result, when engaging in projects, where both user groups interact with each other, researchers and practitioners should put close attention to including both groups as early as possible in the design and development of Visual Analytics systems.

5.3 Limitations

We acknowledge that this thesis has some limitations, which we will address in the following. First of all, it is important to mention that in each of our provided Visual Analytics systems, we gave a clear outline of limitations and how we aimed to address them in future design studies. As a result, we overcame many of the limitations of our systems in future studies. For example, in the work about VIMA [30], we pointed out that domain experts wished to better comprehend the decision-making process of our used random forest. We overcame this limitation with the system RfX [35] in a follow-up design study project.

Probably the most important limitation of our work is the small sample size of our study participant groups, which ranged from four to seven study participants. However, Visual Analytics systems are generally designed to tackle very specific domain problems, which one a hand of users can address. Hence, it is rather unusual to build solutions for specific domain problems, where more than a couple of experts are able to evaluate the proposed approaches as demonstrated by other Visual Analytics systems [9, 14, 47, 70, 113]. Nevertheless, we aimed to party addressing this limitation by introducing more quantitative evaluation metrics to our publications. For example, in the case of VIMA [30], we compared the performance of a random forest trained with labels from VIMA to established testing procedures. In the case of IRVINE [28], we evaluated how labeling speed improved after the system introduction. Regarding RfX [35], we compared how derived rule sets from system interactions performed compared to a random forest and decision tree. In the case of ManEx [33], we evaluated how measurement relations led to a reduction in scraped parts. Nevertheless, other quantitative evaluations, such as lab experiments over a large period of system usage, would add further validity to our research, which is why we aim to continue evaluating our systems in the future.

Furthermore, our theoretical models [29, 29, 27] we also developed following a qualitative research design, for example by performing case studies [136]. This is also due to the novelty of the addressed topic of creating, formalizing, and transferring knowledge with Visual Analytics in industrial manufacturing processes. As a result, we regard our developed theoretical models as an exploratory step toward understanding how knowledge is created, formalized, and transferred with Visual Analytics in manufacturing settings and plan to further extend and validate our theoretical models in future Visual Analytics systems in our application domain.

5.4 Conclusion and Future Work

The work developed during this thesis comprises the investigation of the creation, formalization, and transfer of domain knowledge through Visual Analytics systems in industrial manufacturing settings. Throughout this thesis, we addressed several research gaps regarding the design and development of Visual Analytic systems in manufacturing settings, knowledge creation, formalization, and transfer through Visual Analytics, as well as abstractions and conceptualizations of industrial manufacturing settings. In this regard, we designed, developed, and evaluated five Visual Analytics systems with the goal to create, formalize, and transfer knowledge to improve industrial manufacturing processes. Our developed Visual Analytics systems go beyond traditional business intelligence systems or purely machine learning-based approaches, which aim to improve manufacturing processes.

In the case of business intelligence, we encountered the problem that expert feedback cannot be stored in a system's underlying data structure. However, by leveraging Visual Analytics systems, we showed that knowledge can be formalized via rules, labels, annotations, or models and fed back to the system to introduce human-in-the-loop cycles where not only the system but also the cognition capabilities of human experts evolve over time.

In the case of purely machine learning approaches, we experienced that in many cases such approaches are hard to introduce in industrial manufacturing processes. This was mainly because a high degree of domain knowledge is needed to train such models, which in many cases is particularly hard to access. In this regard, we demonstrated how Visual Analytics systems can support the creation, formalization, and transfer of domain knowledge and how this knowledge can be leveraged in the context of machine learning model building.

In addition to that, we derived two theoretical models, which help to contextualize Visual Analytics systems and formalized knowledge through the Visual Analytics systems in light of organizational knowledge management. These models help to understand how knowledge is created, formalized, and transferred via Visual Analytics. Furthermore, they show how labels - as explicit knowledge products - are created and flow through organizations to be shared among different stakeholder groups.

Finally, we outlined several abstractions of common domain problems, tasks, and data in industrial manufacturing scenarios. For example, we found that an important often addressed domain problem is that experts find it difficult to analyze sensor data of produced parts and their sub-components along with several connected manufacturing stations. One example regarding the abstractions of tasks is the specific labeling of sensor measurements and annotating the cause of found data in the recorded raw data. In terms of data abstractions, we showed that spectrograms - as complex data structures in the field of acoustic testing - can be reduced to three-dimensional data spaces.

That said, we provided various examples of how Visual Analytics supports the creation, formalization, and transfer of expert knowledge. Furthermore, we gave examples of how to design, develop, and evaluate Visual Analytics systems in industrial manufacturing scenarios and illustrated challenges and pitfalls, which might help future practitioners and researchers in engaging in similar problem domains. These findings were supported by our theoretical models, which help to contextualize the knowledge creation, formalization, and transfer via Visual Analytics in light of organizational knowledge management.

Although much was achieved during this thesis, we are several ideas that deserve to be further investigated in future work.

First, we only analyzed sources of externalized knowledge, such as features, labels, or annotations only in the context of one organization and its organizational units. However, more studies need to be carried out about how such explicit knowledge products affect organizational units from multiple organizations. For example, it might be of interest how labeled parts of an industrial manufacturing company can be of value for its supplier or vice versa. In this regard, it is also important to understand how Visual Analytics systems must be designed to improve knowledge creation and transfer between organizations.

Second, we experienced in our research work that the formalization of domain knowledge is a time-consuming and challenging task, which requires lots of experience and understanding of the domain problem. Nevertheless, explicit knowledge has a large impact on the work of multiple organizational stakeholders, such as data experts. Hence, another interesting research direction is to investigate how organizations can design incentives to increase the labeling efforts of domain experts. In this regard, organizations might be interested in considering whether it is important to introduce the role of a "*data labeler*". This job role could include experts with knowledge in the relevant problem domain, such as the acoustic testing of electrical engines as well as a good understanding of data analysis problems, such as model training.

Third, regarding the design of Visual Analytics systems, we experimented with multiple different algorithms to address our relevant domain problems, for example, computation of residual values for sensor measurements or the deviation of sensor recordings to fixed measurement parameters. However, our final Visual Analytics systems are mostly based on single algorithms. As a result, another possible research direction is to analyze how the best algorithms for certain domain problems can be selected during the interactive use of a Visual Analytics system. For example, for the identification of anomalies in acoustic testing, it could be more appropriate to compute residual values while in the case of electrical tests the deviation of a sensor recording to fixed measurement parameters can be a better solution.

Fourth, we already carried out exploratory work on how a design study methodology can serve to address a multi-stakeholder user group. Nevertheless, the original design study methodology by Sedlmair et al. [105] is designed to account for the interaction between a single stakeholder group, such as engineering experts with a Visual Analytics systems. As a result, we propose to further investigate how to enhance the design study methodology to account for different stakeholder groups, such as engineering and data experts.

Fifth, we experienced that the analysis of manufacturing data requires a clear and holistic big picture of manufacturing processes. For example, it is important to know, how many parts were produced at a certain time at which station and what measurements were recorded for each part at each station. To tackle this issue we experimented with knowledge graph-based approaches to add another semantic layer on top of the manufacturing process. However, we found that the holistic information about large-scale industrial manufacturing processes is scattered and fragmented across organizations. Hence, we propose to further investigate how to automatically process information from such data sources and inject the resulting findings into a knowledge graph. In this regard, it is also important to know what information is missing and where to find it. In this regard, we believe that a complete knowledge graph about assembly lines or even a whole manufacturing process would drastically improve data-driven analyses of manufacturing data.

Part II

Part B - Publications of the Dissertation

Chapter 6

Publications (P)

6.1 P1: VIMA: Modeling and Visualization of High Dimensional Machine Sensor Data Leveraging Multiple Sources of Domain Knowledge

Full reference: *J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In Visualization in Data Science at IEEE VIS, 09 2020*

VIMA: Modeling and Visualization of High Dimensional Machine Sensor Data Leveraging Multiple Sources of Domain Knowledge

Joscha Eirich*
University of Bamberg, BMW Group

Dominik Jäckle†
BMW Group

Tobias Schreck‡
Graz University of Technology

Jakob Bonart§
Fraunhofer IWU, BMW Group

Oliver Posegga ¶
University of Bamberg

Kai Fischbach ‖
University of Bamberg

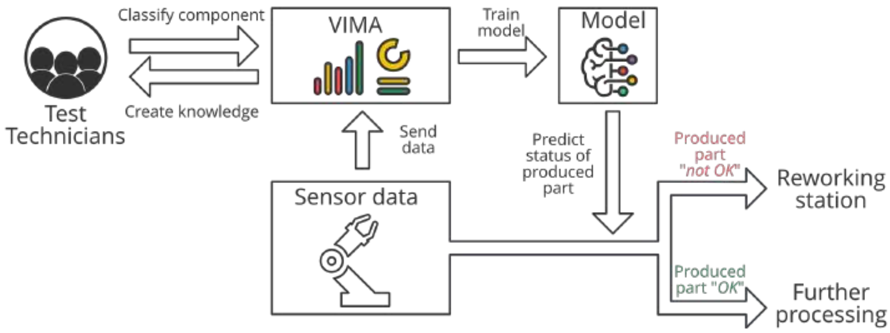


Figure 1: The process of building machine learning models in a real-time manufacturing environment. Machine-sensor data is sent to our VA (Visual Analytics) application *VIMA (Virtual interactive manufacturing assistant)*, where it supports multiple test technicians in their analysis of produced parts. *VIMA* further enables the labeling of parts, which are fed into a machine learning model. The model classifies the status of a produced part based on the machine-sensor data into *not ok (NOK)* and *OK*. In the case of a *NOK* part, it is sent to a specific reworking station for a detailed in-depth analysis. *OK* parts are transferred to the next assembling station.

Abstract

The highly integrated design of the electrified powertrain creates new challenges in the holistic testing of high-quality standards. Particularly test technicians face the challenge, that lots of machine-sensor data is recorded during these tests that needs to be analyzed. We present VIMA, a VA system that processes high dimensional machine-sensor data to support test technicians with these analyses. VIMA makes use of the concept of interactive labeling to train machine learning models and the process model of knowledge creation in visual analytics to create new knowledge through the interaction with the system. Its usefulness is demonstrated in a qualitative user study with four test technicians. Results indicate that through VIMA, previously undetected abnormal parts, could be identified. Additionally, a model trained with labels generated through VIMA, was deployed on a test station, that outperforms the current testing procedure, in detecting *increased backlashes* and improved the test benches output by 15%.

Keywords: Visual analytics, machine learning, knowledge creation, interactive labeling, anomaly detection.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces (GUI); User-centered design

*e-mail: joscha.eirich@uni-bamberg.de
†e-mail: dominik.jaeckle@bmw.de
‡e-mail: tobias.schreck@cgv.tugraz.at
§e-mail: jakob.bonart@bmw.de
¶e-mail: oliver.posegga@uni-bamberg.de
‖e-mail: kai.fischbach@uni-bamberg.de

1 Introduction

The transformation towards the production of fully electrical vehicles, which additionally is pressured by governments across the globe [21], remains a challenge for car manufacturers [14]. Regarding this transformation, the two technologies, *electrical energy storage systems* and *electric engines* build the core of electric vehicles. To ensure their high-quality serial manufacturing, extensive testing steps need to be performed during their assembly. Therefore, automotive companies make huge efforts in the development of so-called test benches [33, 39]. To save costs and to optimize the manufacturing process their task is to test produced parts to exclude them if they do not fit the prescribed quality criteria [30]. Additionally, the trend towards the digitization of many industries (often referred to as Industry 4.0) also enables new generation test benches to generate large amounts of data by multiple sensor recordings, with the promise of creating new usage opportunities [38].

However, test technicians, who are responsible for the development of test benches, face the major challenge that tests for these technologies are just to be developed. Established testing procedures, such as benchmarking single measurements of machine-sensors [17], do not meet the new requirements for the holistic testing of highly integrated powertrains. Machine learning (ML) provides the necessary means to improve product quality by processing huge amounts of data of multiple machine-sensors and exclude produced parts that do not meet quality requirements [25, 40]. Test technicians are required to analyze and label machine-sensor data to produce the necessary test data for respective ML models. However, the successful application of ML methods to production processes faces many challenges in practice, including problems of data selection, cleaning and preprocessing, feature engineering, model selection, and provision of sufficient appropriately labeled training data, to name a few [8]. As a first step, anomalies in the recorded measurements of machine-sensors provide a foundation to train initial ML models that can be improved continuously [2], whereby an anomaly represents a high deviation from of a measurement value in comparison to the typical distribution of measurement values (e.g. single values or time series). Due to the high amount of measurements for each produced part and the high dimensionality of their underlying feature space, the analysis of machine sensor data can often be overwhelming, which results in a pressing need for automated support [38] for the test technicians. This problem leads to the following research question: *How can a test technician be supported in the identification of abnormal vehicle parts based on huge amounts of machine-sensor data?*

To address the research question, we present VIMA (Virtual interactive manufacturing assistant), a VA system, that was developed in close cooperation with the car manufacturer BMW. The goal of VIMA is a first pragmatic attempt to find a solution on how to support test technicians in their analysis of huge quantities of machine data to improve manufacturing processes. Our contributions are as follows: (1) VIMA – an interactive system for the processing and exploration of machine-sensor data designed in cooperation with two test technicians; (2) The application of the process models for *Visual-Interactive-Labeling* and *knowledge creation in VA* in a real-world manufacturing setting; (3) A qualitative user study together with four test technicians for *electrical engines* and a resulting ML model that outperforms a currently applied testing procedure.

2 Related Work

In this section, we briefly describe the state of the art in detecting anomalies in machine-sensor data (Section 2.1) followed by the application of visual analytics in the industrial sector (Section 2.2) and a motivation to make use of the process models of knowledge creation in VA [32] and interactive labeling [7] (Section 2.3).

2.1 Anomaly Detection with Machine Sensor data

Through the connection of multiple machines and the analysis of the recorded sensor data, new opportunities emerge to improve process and product quality in the manufacturing sector [25]. Since the rise of industry 4.0, a vast number of studies were conducted on the analysis of machine sensor data [1,15,17,20,22,28]. To name a few: Hubert et al. [17] use *operational model analysis* for condition monitoring in gear vibrations; Eger et al. [15] use *correlation analysis methods* in multi-stage production system for reaching zero-defect manufacturing; Ahmad et al. [1] further develop methods of *hierarchical temporal memory* for real-time anomaly detection in stream data. All of these approaches have in common, that they operate in an experimental setting, where either very few parts were measured [17,22,28] or created models relied on simulated data [15,20]. Thus, the created approaches are not able to tackle the issue of measurement distributions in a real serial manufacturing environment. During this research, we build on the findings of previous research by applying ML techniques to support domain experts in their analysis of machine sensor data within a real-world setting.

2.2 Visual Analytics for Industrial Applications

Recently, a survey on applications of visual analytics in the manufacturing sector has been published [44]. It acknowledges the need for diverse customized solutions and a demand for visual analytics, to facilitate the analysis of high dimensional multivariate machine sensor data. Some examples are available, on how to support analytical reasoning of test technicians during testing steps [3,24,27] and others on finding anomalies in multivariate time series [18,41]. Since testing phases contain the most data abundant processes in manufacturing processes, an often reported issue of test technicians is to understand data spaces over three dimensions [42,43]. To tackle this issue, Pajer et al. [27] developed a novel system for the exploration of weight spaces. It facilitates the quantitative analysis of internal relations between data instances to explore which instances are most relevant for further analyses. Considering the analysis of multivariate time series of machine sensor data Janetzko et al. [18] propose a similarity and model-based visualization of anomalies with several visualization techniques, such as recursive patterns or line charts. The notion of trust and system adoption remains also an often reported topic, when applying VA in the industrial sector [14,38,44]. All studies reveal a pressing need to include the users of VA systems in the manufacturing sector as early as possible in the development process [2] to facilitate sense-making and enhance problem-solving capabilities. We built on their findings, by developing a generic VA system, that can be adapted individually to the needs of multiple test technicians. As suggested by the studies [2,18,38], we included the user groups of VIMA from the beginning on into our development process, so we can ensure the adoption of the system.

2.3 Interactive Labeling

On the one hand, test technicians face the major challenge that tests for the newly developed powertrain components are just to be developed. Thus, they cannot rely on their experience solely. On the other hand, quality classification models fail to confidently provide reliable results due to the lack of data. To bring the best of both worlds together, Sacha et al. [32] propose to put the user into the loop to generate new, valuable knowledge through exploration and verification. In this process exclusive domain knowledge is leveraged and assisted classification via interactive machine learning enabled [31]. In this context, Amershi et al. [2] stresses the importance of accounting user behavior, to make interactive ML as efficient as possible. In the case of missing labels to build sophisticated ML classifiers, Bernard et al. [7] propose the concept of *Visual-Interactive-Labeling* (VIAL) as a means to visually label yet unknown data in an interactive, explorative setup. VIAL thus bridges the gap between active learning [34] and advanced visualization concepts. Grounded on the process of knowledge creation in VA [32], we describe in Section 5.2 how domain knowledge of multiple test technicians can be leveraged within a manufacturing environment. Additionally, we make use of the process of VIAL in Section 5.3 to demonstrate how it helps to interactively train ML models and evaluate the results on a real test bench for electrical engines in Section 6.2.

3 Background and Case Description

In this section, we describe the actual use case (Section 3.1) and our planned solution for the research problem (Section 3.2).

3.1 Quality Control in Highly Automated Manufacturing

We conducted our work for the automotive company BMW, in particular, its units for the manufacturing of the electrical powertrain. The study was carried out to optimize internal testing procedures of existing test benches for new generation electrical vehicles. During the manufacturing of the electrical powertrain, all of its assembled parts are tested at multiple stations.

A single test of a produced part often lasts only a few seconds. Thus, test benches are upon all optimized to minimize false negatives, meaning that they need to detect parts that contain a functional error. Hereafter, these parts are referred to as "*not OK (NOK)*" parts. Generally, *NOK* rates are very low in highly automated manufacturing environments – often less than 1% – resulting in few labels that can be used to build ML models. Furthermore, the optimization towards low false negative rates – often by lowering thresholds – comes with an increased false positives rate. In this context, false positives can be referred to as tested parts that are classified as *NOK* by a test bench, but are in fact "*OK*". They are often a result of a faulty test run. Even though they do not affect the quality of a part, they have a negative impact on cyclic times, where they need to be tested multiple times until they turn out to be *OK*. For this research, we analyzed data from two of BMW's test benches, one for *electrical engines* and one for *electrical energy storage systems*.

All measurements are recorded by multiple sensors resulting in multidimensional time-series data. Throughout both testing procedures, test technicians observe the tests and are responsible for the condition and performance of the tested parts. During the testing of a part, test technicians are monitoring a few sensors for threshold violations, which they select manually according to their given domain knowledge. However, for the recently developed, highly integrated design of electrical powertrains, there is no to little knowledge on all sensors and the interplay between them. In fact, Schnigg et a. [38] reported in a recent study, that the interplay between the measurements of distinct sensors is an important contributor to the overall quality of a produced part. Therefore, our work is motivated by making use of time-series data of multiple measurements of sensor data and to discover new ways on how to support test technicians with the analysis of these relations. The main goal here lies in gaining new knowledge and built ML solutions to increase the overall quality of produced parts (*false negatives*) without a negative impact on cyclic times (*false positives*). During this research we will mainly use examples from test technicians of *electrical engines*, because the qualitative evaluation was carried out with them and a final ML model deployed on a test bench for *electrical engines*.

3.2 Planned Solution

The planned process model of how we intend to support test technicians with the analysis of their data and training ML models with VIMA is outlined in Figure 1. By using machine-sensor data as an input, VIMA provides an explorative environment that supports the efficient identification of abnormal parts. Through the interaction with VIMA test technicians gain knowledge, since big amounts of unstructured machine data are visualized to them in a meaningful way. When a test technician identifies an apparently abnormal part, he can assign a label to it, which triggers the continuous improvement of the system's underlying ML model. The model then makes recommendations according to the given input further facilitating the identification of abnormal parts. As soon as the model is trained and validated, it is deployed to the test bench, where it then makes real-time predictions. *NOK* parts are then sent to a specific reworking station for an in-depth analysis. Apparently normal parts are further processed by the next station. The model predictions are validated by the test technicians through the continuous analysis of the machine sensor data supported by VIMA. If the test technician disagrees with the model prediction, the model receives a penalizing feedback. When the technician agrees with the suggestion, the model receives positive feedback.

4 Data Processing Pipeline

As a first step to setup VIMA properly in an industrial environment, we developed a data processing pipeline that normalizes machine sensor data (Section 4.1) and extracts the relevant input data (Section 4.2) both for the model and the visualization. The processing pipeline also reduces the dimensionality of the feature space to facilitate the analysis of the data (Section 4.3).

4.1 Normalization

Time series are generated by multiple different sensors with different scales; this makes it difficult to compare them directly with each other. Additionally, during interviews with our test technicians, it was suggested that we need to focus on the changes of patterns rather than numerical values. Thus, we first reconstruct each measurement of the time series through normalization. Here, for each measurement we built a baseline curve by taking the mean of the same measurement for all parts μ and accordingly its standard deviation σ. Then, the residual values of each measurement $X = (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ are calculated as $X_{res} = (0, y_{res1}), (1, y_{res2}), ..., (x_n, y_{res_n})$, where

$$y_{res_i} = \frac{y_i - \mu_i}{\sigma_i}$$

The transformation from real measurement values to its residual values is shown in Figure 2.

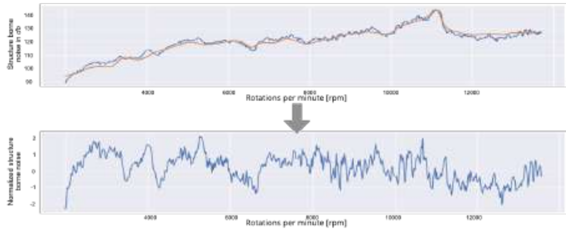


Figure 2: The normalization of a measurement. In the upper plot the mean baseline curve μ_i is colored in orange and the real values in blue. The curve in the lower plot represents the residual values.

4.2 Feature Selection

After all measurements are normalized, we set up the data processing pipeline, where we construct the feature matrix as input data both for VIMA and its underlying model. This process is outlined in Figure 3. First, in (a), for each produced part *(P)*, k measurements *(M)* are selected. One measurement, for example, is the structure-borne noise in decibel (dB) to measure the unbalance of an engine and another the torque of an engine over a certain rotational speed per minute. Next, in (b), we sat down with each test technician both from *electrical energy storage systems* and *electrical engines* to determine the most important features of the measurements. For example in the case of *electrical engines*, all measurements that were recorded with a rotational speed lower than 3000 rpm were considered as irrelevant, where the remaining values were separated into five equal segments *(S)*, in (c). We then provide the possibility for the test technicians to extract the generic features *(F)*, *minimum, maximum, mean, standard deviation* in (d). Whereas for *electrical energy storage systems*, the test technicians choose to select all possible features, for *electrical engines* only the features *maximum, mean*, and *standard deviation* were chosen. Even though we did experiment with different features, such as autocorrelations [16] or indexes of local maxima, these standard features are well-established features to conduct machine learning with time-series [8]. Furthermore, they are most understood by our test technicians and were adopted very well. Thus, we applied the processing pipelines to the measurement values of both test technician groups resulting in the feature vectors for each part with:

$$dim(\vec{P}_n) = M_k \cdot S_i \cdot F_j$$

All machine-sensor data that we use has in common, that it consists of only numeric time series (for instance, engine noise over rotational speed) and therefore can be processed the same way. Our processing pipeline can be easily customized, for example by selecting different segments or features, it can be used for any kind of time series data. Thus, it also can be generalized to use cases outside of a manufacturing context. Examples include sales data, where products (P) in different regions (R) can be divided into i segments (S) and j features (F).

4.3 Dimensionality Reduction

The purpose of dimensionality reduction lies both in the improvement of learning algorithms [19] and the interpretability of a high dimensional data space [43]. Here, the main benefits are visual depiction of the data so that users have a means to explore the high dimensional space. In this regard, we use a principal component analysis *(PCA)*, which is a well-established method for feature extraction and dimensionality reduction [37]. Since *PCA*, is an often

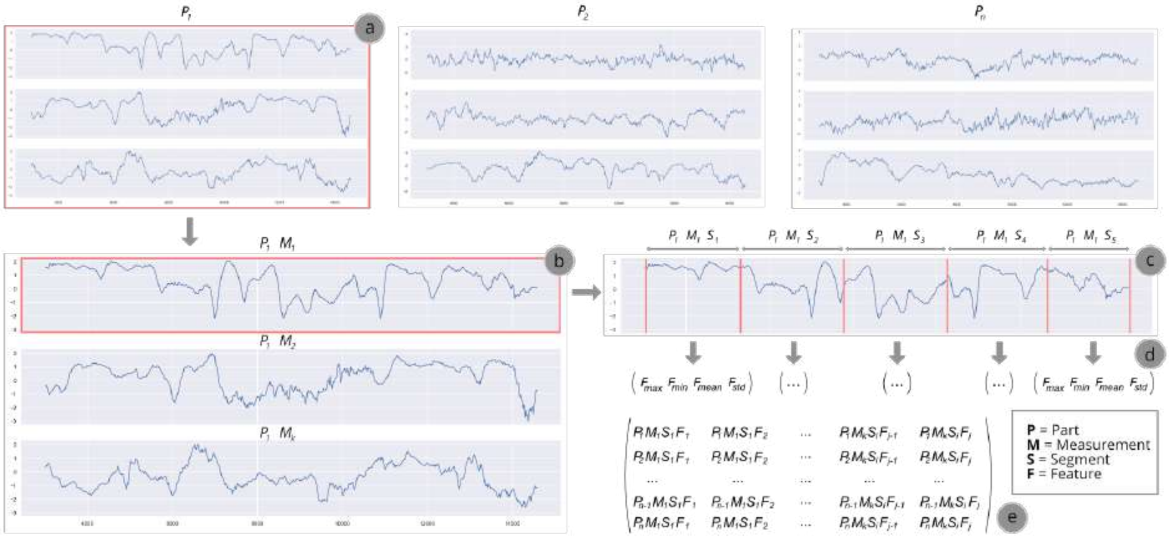


Figure 3: Feature engineering process to generate residual value matrix with data from *electrical engines*. First in (a), for all parts *(P)*, each single *P* is selected. Then in (b), for each *P*, each measurement *(M)* is divided in (c) into segments *(S)*. For each *S* in (d), the values: *maximum, minimum, mean, and standard deviation* represent our features *(F)*. This results in a feature input matrix in (e).

applied technique in the analysis of linear dependent time series [19], we consider it as an adequate application for our use case. We are aware of the fact that data visualizations in 3D might contain occlusion and perspective distortion effects [11]. It also poses challenges for user navigation, and precise selection operations due to increased interaction demands. In general, it is difficult to say whether 2D or 3D is more effective, as it may depend on the application, tasks, data, display, and interaction environments [23].

Even though the involved test technicians were familiar both with 2D and 3D visualizations, they explicitly opted for a 3D representation instead of three 2D visualizations. This was due to the fact, that they found it confusing to interact with three separate 2D plots simultaneously instead of one single 3D plot. Moreover, a third principal component (*PC*) added additional variance both for *electrical engines* and *electrical energy storage systems* as shown in Table 1. We thus decided to use a 3D instead of a 2D representation. However, a switch to 2D representations is no conceptual problem.

	PC1	PC2	PC3
Electrical engine	0.1215	0.0886	0.0706
Electrical energy storage system	0.3825	0.2386	0.1184

Table 1: Explained variance of principal components (PC) for both use cases.

5 VIMA

In the following section, we discuss the design of VIMA. First, we describe the methodology that was used to design the system (Section 5.1), followed by a discussion of the main parts of our approach: The role of domain knowledge in VIMA (Section 5.2), the interactive model training process (Section 2.3), and the design of the visual interface (Section 5.4).

5.1 System Design

The design of VIMA is based on the *Design Triangle* by Miksch and Aigner [26], which suggests considering the *users*, the *data*, and the *tasks* as main factors when designing and implementing a VA application for time-oriented data. To fully comprehend the problem domain, we developed the system in close cooperation with two test technicians from BMW; one was focusing on the development of test procedures for *high voltage batteries* and the other one on procedures for *electrical engines*. Both had a mechanical engineering background and were not familiar with data science methods. They guided us in the development of VIMA iteratively and provided us with feedback about relevant tasks. We used their domain knowledge to assemble our design triangle as follows:

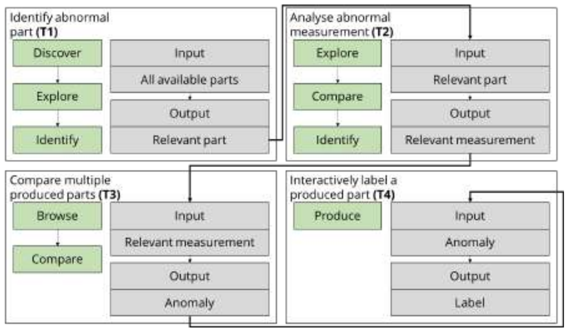


Figure 4: The identified tasks according to [9].

Users: VIMA targets test technicians in the manufacturing of electrical vehicles. They develop testing procedures for multiple types of parts and are responsible for the development of test benches. Due to their extensive domain knowledge, only they are able to provide context to the data that is processed and visualized through VIMA and have the ability to label produced parts accordingly. Furthermore, through the interaction with VIMA, they can create new knowledge since the data of their test stations is presented to them for the first time considering the relations between multiple parts.

Data: The data are measurements from machine-sensors, which are generated by test benches. Each sensor measurement comprises multivariate time-series data. An example of such measurements is the structure-borne noise to test the physical behavior of an *electrical engine*. To ensure, that data from different test benches can equally

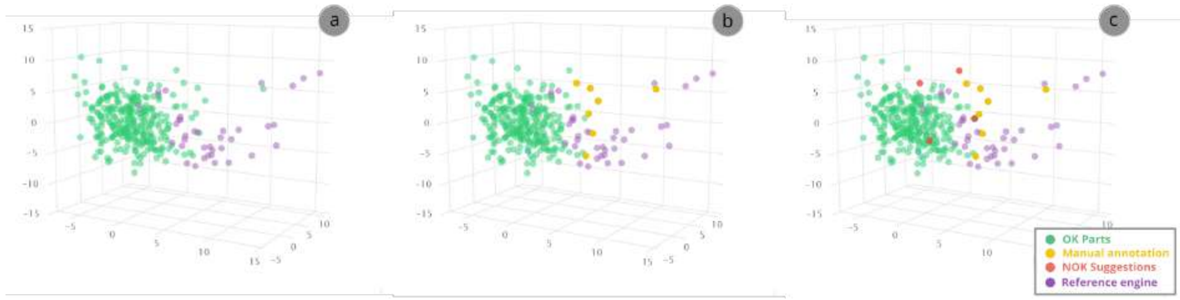


Figure 5: Exemplary interactive model training of VIMA. In (a) only reference engines are marked as purple. Then the test technicians select manually engines, that are located close to purple ones or at the edge of the green main cluster in (b), which are marked yellow. With these labels, a first baseline model is trained in (c) where red dots indicate a prediction for an *NOK* part.

be visualized with VIMA we use our preprocessing pipeline as outlined in Section 4. In this step the data is normalized (Section 4.1), features are extracted (Section 4.2) and dimensions are reduced to three principal components (Section 4.3).

Tasks: We focus on user tasks related to analyze large amounts of machine-sensor data. To derive the tasks we followed the task taxonomy of Brehmer and Munzner [9] and conducted multiple interviews with the test technicians. We chose this taxonomy because it forms the bridge between high and low-level tasks [9], which is particularly helpful to embed them into the daily routines of the test technicians. Figure 4 visualizes the tasks that we identified, which are the following:

T1: *Identify abnormal parts*. Users monitor parts on multiple hierarchical levels, i.e., a produced part with the most abnormal features. For an in-depth analysis, a relevant part has to be discovered, explored, and identified by the user.

T2: *Analyse abnormal measurement*. Since each part contains multiple measurements of machine sensors when a relevant part is selected the user also investigates how the explicit time series of a measurement contribute to its anomaly. After the user selected a part, he/she can browse for most deviating measurements by analyzing each measurement of the selected part individually.

T3: *Compare multiple produced parts*. The user compares the previously identified measurement to multiple reference or apparently normal parts. Thus, his/her assumptions about the initially selected part can be validated.

T4: *Interactively label a produced part*. When the user finishes his/her analysis he/she produces a label for the selected part. A part can be labeled as *OK* or *NOK*, which is then fed into the system's underlying ML model.

5.2 Incorporation of Domain Knowledge

To make use of as much domain knowledge as possible we integrated the test technicians from an early stage into the development process [2]. During the interaction with VIMA, all phases of Sachas' et al. [32] process model of knowledge creation *(exploration, verification, and knowledge creation)* are addressed [32].

With the initial input data, the test technician can select a specific part for an in-depth analysis. During the *exploration phase* [32], the test technician interacts with VIMA and changes the visualizations to better understand the behavior of an analyzed part. This is achieved by identifying abnormal parts with 3D plots and heatmaps, through evaluating the cause for the anomaly with lists and line charts, and by comparing them with other engines. All views are described in detail in Section 5.4. Each interaction with a single visualization causes an adaption of another visualization. Selecting for example a part in the 3D plot results in the adaption of the lists and line charts, which give the test technician an immediate hint on how to continue his/her analysis. After his/her analysis of the part he/she generates findings, that, when considering previous domain knowledge may become an insight, often referred to as the "Ah Ha" moment [12]. In this *verification phase* [32], an insight can be quite small, such as realizing that two measurements, that were considered as independent, are actually related to each other. Additionally, Sacha et al. [32] argue that an insight might also result in the formulation of hypotheses, that must be tested accordingly. An hypothetical example can be: *"A high structure-borne noise decreases the overall quality of the part"*. Whenever the test technician trusts the insight, and internalizes it, he/she gains new knowledge during the *knowledge generation phase* [32]. A hypothetical example of this knowledge is the discovery of a hidden pattern in the measurements that always leads to decreased quality of a part during its manufacturing. Nevertheless, hypothesis need to be properly examined to avoid false conclusions.

When the test technician finishes his/her analysis, he/she can use the previously generated knowledge to assign a label to the selected part. Here, labels represent the aggregation of complex information about an observed data instance [6], that only can be provided by test technicians. Before an initial model is trained, the pure visualization of the data allows the user to select the most abnormal parts and create labels for them. After an initial set of labels is available, a first baseline model is created to suggest possible classes to the test technician for further analysis.

5.3 Labeling and Model Training

VIMAs classifier, is built with a standard implementation of Breiman's random forest [10]. During the interactive labeling through VIMA, the three goals *(G)* of VIAL *(gained domain knowledge* (G1)*, labeled data* (G2)*, and trained learning models* (G3)) [7] are addressed. As input features, we use the feature matrix described in Section 4.2. Figure 5 provides an overview of the interactive labeling process for *electrical engines*.

As a first step, we need to check whether reference parts are available or not. A reference part is a part, that is damaged by design to test certain possible error outcomes. In the case of an *electrical engine* a reference part can be an engine, with a rough surface on the spur gears. This reference engine behaves differently according to its structure-borne noise compared to engines that contain spur gears with clean surfaces. Even though reference parts are important to detect certain errors, the possibility of undetected errors always remains. Thus, VIMA does not only enable a visual separation of reference and regular parts but also of parts that contain abnormal measurements. Reference parts are marked purple and the rest green in Figure 5 in (a). Since in a manufacturing setting a high percentage of parts is generally *OK*, the majority of the parts belong to a single cluster, which is clearly visible in (a). Hence, if no reference part is present, the visualization still provides sufficient information about, which parts contain the most abnormal measurements and should

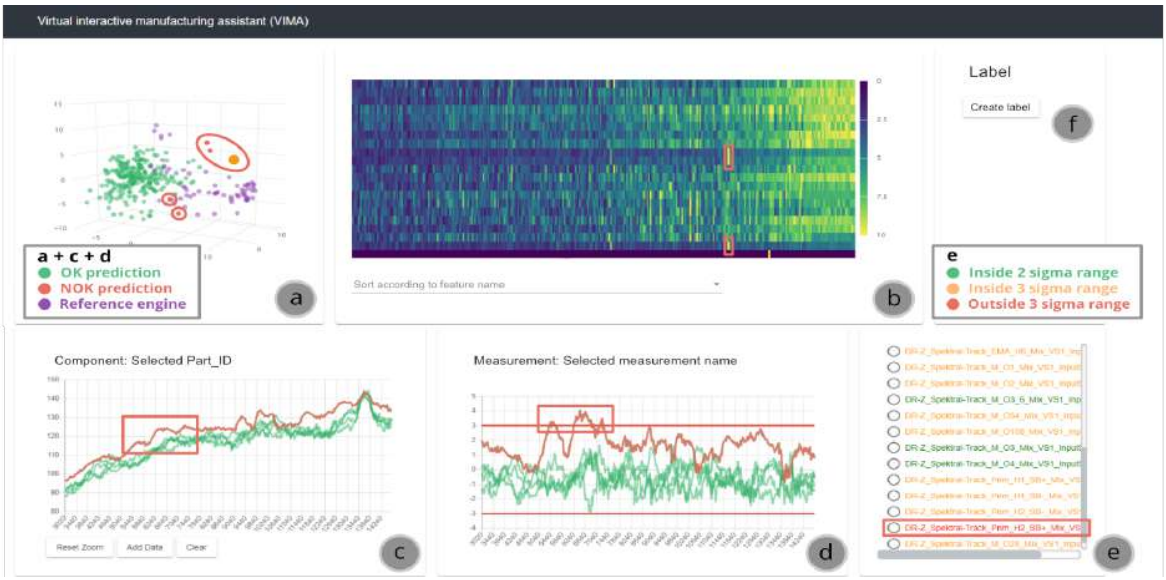


Figure 6: Screenshot of *VIMA* with data from *electrical engines*. The two views at the top (a,b) provide an overview of the behavior of produced engines, where (a) is the projection of features in 3D via principal components; (b) is a sorted heatmap of all normalized features according to the first component. The views at the bottom (c,d) represent the real measurement (c) and its normalized residuals (d), with corresponding colors to (a); (e) provides all available measurements for a selected engine; (f) provides the user with the possibility to label the selected part.

thus be labeled as *NOK*. Here, parts that lie outside or at the edge of the main cluster of apparently *OK* parts are more appropriate candidates for a deeper analysis than parts that are directly inside the cluster. In (a) the test technician gets an initial overview, which parts are outside the green cluster and/or close to a purple dot. By selecting one of these green engines, new knowledge *(G1)* is created through the analysis of the part as described in Section 5.2.

After the test technician finished his/her analysis, he/she can assign a label to a part. In our case in Figure 5, all of these manually created labels are colored in yellow for demonstration purposes in (b). Both the reference parts and the manually assigned labels build the core set of labels to train an initial model and thus reflect *(G2)*. Based on these labels, the model makes further suggestions which other parts might contain anomalies. We ensure a deterministic implementation of our random forest by defining a threshold that maximes the number of true positives and true negatives. Based on this threshold, model predictions for probable NOK parts are shown in (c) as red dots. As soon as this baseline model is trained, the test technician now either needs to classify model predictions as *wrong* or validate them as *true*. This feedback is then constantly sent to the model for retraining purposes, which results in trained and learning models *(G3)*. The first model is trained on historical data that is loaded into VIMA. It is then retrained every time a new label is added by a test technician. To ensure its deterministic implementation, the threshold also has to be adapted constantly. To avoid pitfalls such as overfitting or wrong classifications, the model must be validated continuously by the test technicians through VIMA.

5.4 Visualization

A screenshot of VIMA, with data from *electrical engines* is shown in Figure 6. It consists of six components, which are labeled from a-f. VIMA is designed according to Schneiderman's visualization mantra *"Overview first* (**T1**) *- then details on demand"* **(T2, T3)** [35], with the goal to meet the requirements of the tasks outlined in Section 5.1.

During multiple analyses of the data, the test technician created labels, that are used to train and improve our model, as outlined in Section 5.3 and to enhance the visual analysis of produced parts, as outlined in Section 5.2. Additionally, all red circles and squares in Figure 6 indicate exemplary manually annotated anomalies for demonstration purposes. In the 3D plot (a) all available parts are presented, where each axis represents one principal component from our data processing pipeline according to Section 4.3. Here, green dots indicate an *OK* and red a *NOK* model prediction. The model is an additional visual help for the user, where instead of finding an interesting green dot, he/she can select a red dot that was suggested by the model. Purple dots represent a reference engine.

When a model is not trained and reference parts are not available, initially all dots are colored as green, where the user has to choose abnormal parts based on their distance to the *OK* cluster. The limitations of 3D visualizations are perspective distortion and occlusion [11]. We are aware of this limitation but acknowledge, that this representation was explicitly requested by the involved test technicians, both from the *high voltage storage battery* and *electrical engines* group. Here we agree with Amershi et al. [2] that systems like VIMA should aim to empower *"end users to create machine-learning-based systems for their own needs and purposes"* [2] and thus put the test technicians above all in charge of the final decision which visualization to use. Nevertheless, to alleviate this issue, the view angle can be adapted dynamically.

The colors red and green are also used in BMWs serial manufacturing environment. Purple was especially requested since it was easy to detect these engines immediately as reference engines. The feature matrix is presented in a heatmap (b). It is initially sorted according to the first principal component, in order to show, that engines that are close in (a), are also close in (b). In the heatmap, we used a blue-yellow colormap, where yellow represents a very abnormal feature and blue a normal one. To detect hidden patterns, the heatmap (b) can be sorted by feature names.

When the test technician either clicks on a part in the 3D plot (a) or the heatmap (b), it is highlighted (see orange circle (a) and yellow mark in the bottom of (b)). In our case, the test technician selects an apparently abnormal part **(T1)** in the 3D plot (a), which is marked with an orange circle in Figure 6. He/She can choose one measurement responsible for the anomaly **(T2)** by clicking on the list in (e) and compare it to the measurements of other apparently normal engines **(T3)**. In the list (e), we color the measurement name green if it is within a two-sigma range compared to all other measurements, orange when it lies between a two- and three-sigma range and red when parts of the values are outside a three-sigma range. We apply the sigma rule, because it is a well established method to detect outliers in manufacturing settings [29] to which the test technicians are already used to. This color-coding based on sigma values helps the test technicians to immediately identify measurements that with high probability were the cause for an abnormal behavior.

When the technician clicks on a measurement **(T2)** both the real measurement values and its residuals are shown in the line charts (c) and (d). Additional red lines in (d) indicate whenever a measurement value lies outside a three-sigma range. We use line charts since they are tailored to the representation the test technicians are most familiar with and are also applied in industrial applications of VA [18]. Especially the line chart in (c), is the same one, that experts use in their daily routines. In turn, the chart in (d) helps especially to immediately verify if a measurement contains values outside a serial distribution, which is a strong indicator for anomalies [17]. The name of the selected part is shown on top at (c) and the name of the selected machine-sensor at the top of (d). After the test technician finishes his analysis he can assign a label to the selected part in (d). The general workflow to use VIMA is executing the task **(T1)** to identify an abnormal part, then within this part find the most abnormal measurement **(T2)**, compare it to other interesting parts **(T3)**, and lastly produce a label **(T4)** to train a baseline model or enhance an existing one. We choose this workflow since it is an appropriate implementation of Schneiderman's visualization mantra [35], where the tasks **T1-T3** are tailored in a way to facilitate knowledge creation according to Sachas' et al. [32] process model and **T4** to interactively create and enhance ML models.

6 Evaluation

In this section we demonstrate the usefulness of VIMA with a qualitative user study (Section 6.1). We then show, how the labels created with VIMA resulted in a model, that managed to outperform an existing testing procedure inside BMW (Section 6.2).

6.1 User Study

In our evaluation, we show how test technicians can be supported in the identification of abnormal vehicle components based on huge amounts of machine sensor data. We choose a qualitative approach by conducting a think-aloud study [36]. A qualitative approach is appropriate, because we follow a general research question without a hypothesis, where we analyze a small sample (few test technicians) in a real-life setting (real manufacturing environment) [13].

Participants: We conducted the study with four test technicians from BMW, responsible for the development of testing procedures for the serial production of electrical engines. They were all male between 23 and 34 years old and had a mean work experience of 5 years and 7 months. Every test technician had a background of mechanical engineering and reported to be not familiar with data science methods.

Data: Data, comprising of multidimensional machine sensor measurements of a test bench in one of BMWs plants, of a period of four weeks (01.04.2020 - 30.04.2020) was loaded into VIMA. The data set comprised of 3,65 GB recorded by a single sensor inside a test bench for *electrical engines*. Due to non-disclosure agreements with BMW either the number of produced engines nor part ids can be named during this study.

Task: The following task was given to each user: *"Identify an abnormal engine according to your point of view"*. The expected execution of the task is: Select a part that is marked red or very close to a purple one in the 3D plot or marked as yellow in the heatmap. Then load an abnormal measurement into the line charts and compare it to apparently normal parts. The expected outcome is an engine that contains an anomaly, where the test technician can decide to label it as *NOK* and send it to the reworking station or as *OK*, whenever it lies within the serial distribution and still meets the quality requirements.

Procedure: All interviews were performed online via Skype or Teams, where each participant analyzed the same data set and executed the above-mentioned task. The interviews were recorded, transcribed, and analyzed by the interviewer post hoc. After the study, we asked the participants to provide negative and positive feedback about VIMA. Each interview lasted between 45-55 minutes.

Results: All participants first looked at the 3D plot (see Figure 6a) and selected an engine that was outside or on the edge of the green cluster. They all noted that the 3D plot was very helpful to get an initial understanding of the selected component. The rotating functionality of the plot as well as the color-coding were particularly reported as very useful. One test technician noted that *"the 3D plot helps me a lot to assess the distance [of abnormal engines] to good engines"*. As soon as an abnormal engine was selected, three participants turned to the heatmap (see Figure 6b) representation, to further analyze it. The blue-yellow color scale was noted as very helpful to efficiently identify a measurement outside the normal distribution. The participant, who did not use the heatmap, reported that the 3D plot already provided him with sufficient information in terms of which component to analyze in more detail. After selecting a specific engine, all participants selected a measurement, which was marked as red in the measurement list (see Figure 6e) and compared it to other normal engines from the green cluster. Here they used the line chart with real measurements (see Figure 6c) and the residuals (see Figure 6d). Three test technicians used the left line chart, because it represents a familiar environment. The residuals were considered as an important representation to see the course of a measurement compared to all other engines (*"I don't need to look at all engines, because here I can see a single engine compared to all others"*) and to validate if an abnormal engine was actually a real anomaly or just within an acceptable deviation of all engines.

All participants successfully identified an abnormal engine within 10-15 minutes and suggested to send the selected engine to the reworking station. Furthermore, they completed the task as expected at the beginning of the study by first selecting an evidently abnormal engine and then comparing its measurements to evidently normal engines. When asked about their expectation of VIMA before the experiment, one participant reported that VIMA *"provides me with a good overview, where I am able to find an abnormal engine very easy compared to the system we are currently using"* and that it was very helpful *" to get a feeling about the behavior of a selected measurement, which is very important"*. The overall feedback for VIMA was very encouraging and indicates that if fulfills the tasks presented in Section 5.1.

6.2 Model Results

Additionally to the user study, we iteratively trained and enhanced a ML model to detect errors in a real-time manufacturing setting. Again, due to nondisclosure agreements with BMW we cannot name absolute but only relative numbers of produced parts. Here, we used the same data, which we used in Section 6.1. As test data we gathered sensor data of two more weeks (01.05.2020 - 15.05.2020), comprising of 1,83 GB disk space. To prevent overlapping data, we recorded the test data directly after the training data.

Two test technicians created the labels with VIMA, as described in Section 5.2. Since, for a real-time manufacturing environment it is important not only to detect a *NOK* successfully but also to know, why this error was produced. Instead of a general *NOK* label, we chose the specific functional error type *"increased backlash (IBL)"*.

The quantity of gear backlash is a combined value of tolerances in gear grinding and the positions of the bearing seats. Therefore, it is an indicator for drifting processes during the manufacturing of gears. If the gear backlash drastically exceeds the specified tolerances it can also lead to more severe faults as elevated noise levels or even tooth breakage.

For our initial model, we used reference parts for *IBL* errors as labels and added manually annotated ones with the help of VIMA. We generated features according to our data processing pipeline outlined in Section 4.2. During the work with VIMA one test technician noted that *"previously we were not able to detect IBL errors with our current software"*. Due to the new visualization provided by VIMA he proposed the hypothesis: "*IBL errors can be explained through sensors that do not measure backlashes but rotor unbalances"*. This implies that sensors that technically are not installed to detect *IBL* errors could be used to detect them, without using the actual sensors developed for the *IBL* test. A successful model, capable of predicting this error would mean, that the actual *IBL* test can be removed from the test bench, resulting in less testing time for each produced engine and thus affect positively the overall cyclic time. We thus decided to test the hypothesis of the test technician to find out if a model could be built with these measurements.

As a model we again make us of random forest [10] and as a benchmark we compared it with the current test for *IBL* errors. To tackle over-fitting we did not split one data set into test and training data but used two separately recorded data sets as described above. Additionally, we applied 10-Fold cross validation to our model training, which is an often used approach in ML to reduce over-fitting [4]. The resulting confusion matrix is shwon in Table 2.

		Real values	
		OK	IBL
Model results	OK	0.9683	0.0053
	IBL	0.0316	0.9946
Current IBL test	OK	0.9500	0.0000
	IBL	0.0500	1.0000

Table 2: Comparison of our model (top) to the current testing process for IBL (bottom).

In comparison to the current test we managed to produce less false positives (0,0316) than the actual *IBL* test (0,0500). In turn, the current test is able to detect all true negatives, where we have a true negative rate of 0,9946. However, the false-negative rate of 0.0053 in our model, is very low and the RocAuc of our model with 0.9838 outperforms the current test that has a RocAuc of 0,9751. Since we managed to train a model, with comparable outcomes to the current *IBL* test, we can remove this test from the test bench resulting in an overall output increase of 15% for the designated test bench.

Thus, the hypothesis created by one test technicians, which was afforded by VIMA, proved to be true and resulted in a direct optimization of the overall manufacturing process.

7 Discussion

In today's highly automated manufacturing settings, large amounts of machine-sensor data [25] can often be overwhelming for the perception of the human mind [38]. Especially the detection of anomalies in an environment, where almost all produced parts turn out to be *OK* and little domain knowledge for new cutting edge technologies is available, remains a challenge.

To address our research question, we propose VIMA, a VA tool, that affords multiple spirals of knowledge creation through continuous interaction between test technicians and the system. This knowledge is created, since VIMA visualizes sensor data in a way that previously was not available for test technicians. By visualizing machine-sensor data both in ways, that test technicians are used to through line charts, and novel ways such as heatmaps or 3D Plots, we managed to present large amounts of time series data in a structured way, where the finding of hidden patterns, and the creation of hypotheses to optimize the overall manufacturing process are all facilitated. Here Sachas' et al. [32] process model of knowledge creation in VA, turned out to be an appropriate model to describe the process of knowledge creation during the interaction with VIMA.

Additionally, we showed that the concept of VIAL [7] is a valuable model to create a continuous stream of labels interactively to train and enhance ML models. Unlike the model-centered concept of active learning, where a model proactively queries for user feedback about selected samples [5], we chose a user-centered approach, where exiting domain knowledge of test technicians is leveraged and new knowledge is created. We thus showed, that through the interaction with VIMA all goals of VIAL (*gained domain knowledge, labeled data, and trained learning models* [7]) are accomplished through VIMA. Knowledge was gained, since hypotheses were made and confirmed by test technicians, labeled data was produced by using VIMA, and models were trained in multiple iterations with these labels.

Even though our solution is a use case specific implementation for BMW, it worked with minor adaptions both for test benches for *electrical engines* and *electrical energy storage systems*. We thus draw the conclusion, that it also can be used for time series data for test benches from any kind of manufacturing context. Nevertheless, applying VIMA to different time series data sets remains subject to future research. The experts also pointed out some limitations of VIMA. In the current implementation, of the systems for *electrical engines*, we only included sensor data of a single test bench. However, sensor data of previous processing steps is considered as a relevant data source by the test technicians. Thus, the usage of further sensor-data will be subject to the second development cycle of VIMA. Currently, a binary classifier of the model distinguishes between *OK* and *IBL* parts. However, other error types would especially help in the 3D plot, where they could be related to different regions of the plot. Thus, we aim to improve the model to such an extent to perform a multi-label classification. Another limitation of the study is that only experts responsible for *electrical engines* were included. However, VIMA can be used for any type of time-series data, when using our data processing pipeline and visualization interface.

8 Conclusion and Future Work

In this paper, we presented VIMA, an exploratory work to create a system that facilitates the analysis of large amounts of machine sensor data in order to support test technicians in their development of cutting edge technologies in the manufacturing of electric vehicles. Unlike many state of the art techniques that operate within a lab setting, our system is applied to a real-world manufacturing environment. Here, grounded on Sachas' et al. [32] process model for knowledge creation, the system leverages domain knowledge and based on Bernards' et al. [6] model of VIAL, ML models can be created interactively and deployed to test benches. A validation through a think-aloud study with four test technicians backed our design choices. Also a model, that was created with features and labels afforded through VIMA outperformed the exiting testing procedure to detect increased backlashes.

Fueled by the positive feedback and model results, future research efforts will build on this foundation by including sensor data from other machine sensors and enhancing the existing ML model. Furthermore, a long term study will evaluate the implications of frequent expert labeling on model performance. In this study we will evaluate if models created with VIMA will improve over time or might converge. In addition, we will evaluate the benefits of different visualizations (e.g. three 2D plots vs. a singe 3D plot) to find out how VIMA can be improved. Finally, the analysis of how the decisions of our model can be properly visualized remains an interesting stream for future research.

REFERENCES

[1] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134–147, 2017.

[2] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, Dec. 2014.

[3] P. Angelelli and H. Hauser. Straightening tubular flow for side-by-side visualization. *IEEE transactions on visualization and computer graphics*, 17:2063–70, 12 2011. doi: 10.1109/TVCG.2011.235

[4] M. Awad and R. Khanna. *Machine Learning and Knowledge Discovery*, pp. 19–38. 01 2015. doi: 10.1007/978-1-4302-5990-9_2

[5] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2017. doi: 10.1109/TVCG.2017.2744818

[6] J. Bernard, M. Zeppelzauer, M. Lehmann, M. Müller, and M. Sedlmair. Towards user-centered active learning algorithms. *Comput. Graph. Forum*, 37:121–132, 2018.

[7] J. Bernard, M. Zeppelzauer, M. Sedlmair, and W. Aigner. VIAL: a unified process for visual interactive labeling. *Vis. Comput.*, 34(9):1189–1207, 2018. doi: 10.1007/s00371-018-1500-3

[8] C. Bishop. *Pattern Recognition and Machine Learning*, vol. 16, pp. 140–155. 01 2006. doi: 10.1117/1.2819119

[9] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, Dec. 2013.

[10] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, Oct. 2001.

[11] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. Extending distortion viewing from 2d to 3d. *IEEE Computer Graphics and Applications*, 17(4):42–51, 1997.

[12] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Comput. Graph. Appl.*, 29(2):14–17, Mar. 2009.

[13] J. W. Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, 2013.

[14] Q. Demlehner and S. Laumer. Shall we use it or not? Explaining the adoption of artificial intelligence for car manufacturing purposes. In *Proceedings of the 28th European Conference on Information Systems (ECIS) 2020*, 2020.

[15] F. Eger, C. Reiff, B. Brantl, M. Colledani, and A. Verl. Correlation analysis methods in multi-stage production systems for reaching zero-defect manufacturing. *Procedia CIRP*, 72:635–640, 2018. doi: 10.1016/j.procir.2018.03.163

[16] B. Fulcher, M. Little, and N. Jones. Highly comparative time-series analysis: The empirical structure of time series and their methods. *Journal of the Royal Society, Interface / the Royal Society*, 10:20130048, 03 2013. doi: 10.1098/rsif.2013.0048

[17] E. Hubert, P. Borghesani, A. Barrau, M. El Badaoui, and R. Randall. Challenging the multiplicative model used for gear vibration. In *Surveillance, Vishno and AVE conferences*. INSA-Lyon, Université de Lyon, Lyon, France, July 2019.

[18] H. Janetzko, F. Stoffel, S. Mittelstädt, and D. Keim. Anomaly detection for visual analytics of power consumption data. *Computers Graphics*, 38:27–37, 02 2014. doi: 10.1016/j.cag.2013.10.006

[19] S. Khalid, T. Khalil, and S. Nasreen. A survey of feature selection and feature extraction techniques in machine learning. pp. 1–8, 08 2014.

[20] D. Kumar, T. Sambharam, S. Kottalgi, P. Mandloi, and O. Kesarkar. Electric vehicle powertrain multiphysics nvh simulation. pp. 490–495, 10 2018. doi: 10.1109/IECON.2018.8595134

[21] M. Lehmann and H. Kuhn. Modeling and analyzing sequence stability in flexible automotive production systems. *Flexible Services and Manufacturing Journal*, 01 2019. doi: 10.1007/s10696-019-09334-x

[22] A. Malhi and R. Gao. Pca-based feature selection scheme for machine defect classification. *Instrumentation and Measurement, IEEE Transactions on*, 53:1517 – 1525, 01 2005. doi: 10.1109/TIM.2004.834070

[23] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger. *Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation*, pp. 25–55. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-030-01388-2_2

[24] K. Matković, D. Gracanin, R. Splechtna, M. Jelovic, B. Stehno, H. Hauser, and W. Purgathofer. Visual analytics for complex engineering systems: Hybrid visual steering of simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 20:1803–1812, 12 2014. doi: 10.1109/TVCG.2014.2346744

[25] K. Mertens. Herausforderungen beim einsatz von datenanalytik für eine ressourceneffiziente produktion. *BHM Berg- und Hüttenmännische Monatshefte*, 01 2019. doi: 10.1007/s00501-018-0803-z

[26] S. Miksch and W. Aigner. A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data. *Computers Graphics, Special Section on Visual Analytics*, 38:286–290, 2014.

[27] S. Pajer, M. Streit, T. Torsney-Weir, F. Spechtenhauser, T. Moller, and H. Piringer. Weightlifter: Visual weight space exploration for multi-criteria decision making. *IEEE Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016. doi: 10.1109/TVCG.2016.2598589

[28] A. Popkov. Usage autocorrelation function in the capacity of indicator shape of the signal in acoustic emission testing of intricate castings. *Journal of Physics: Conference Series*, 671:012047, jan 2016. doi: 10.1088/1742-6596/671/1/012047

[29] F. Pukelsheim. The three sigma rule. *The American Statistician*, 48(2):88–91, 1994.

[30] A. Rassõlkin and V. Vodovozov. A test bench to study propulsion drives of electric vehicles. pp. 275–279, 06 2013. doi: 10.1109/CPE.2013.6601169

[31] D. Sacha, M. Sedlmair, L. Zhang, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. What you see is what you can change: Human-centered machine learning by interactive visualization. *Neurocomputing*, 268:164–175, 2017. doi: 10.1016/j.neucom.2017.01.105

[32] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014.

[33] M. Salah, M. Abu mallouh, M. Youssef, E. Abdelhafez, M. Hamdan, and B. Surgenor. Hybrid vehicular fuel cell/battery powertrain test bench: design, construction, and performance testing. *Transactions of the Institute of Measurement and Control*, 39, 04 2016. doi: 10.1177/0142331216642835

[34] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[35] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, p. 336. IEEE Computer Society, USA, 1996.

[36] M. Someren, Y. Barnard, and J. Sandberg. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses*. 01 1994.

[37] A. Subasi and M. Gürsoy. Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications*, 37:8659–8666, 12 2010. doi: 10.1016/j.eswa.2010.06.065

[38] J. Suschnigg, B. Mutlu, A. Fuchs, V. Sabol, S. Thalmann, and T. Schreck. Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Proc. 3rd Int. Workshop on Big Data Visual Exploration and Analytics EDBT/ICDT 2020*, 2020.

[39] M. Van der Seijs, E. Pasma, D. Klerk, and D. Rixen. *A Comparison of Two Component TPA Approaches for Steering Gear Noise Prediction*, vol. 4, pp. 71–79. 02 2015. doi: 10.1007/978-3-319-15209-7_7

[40] D. Wu, C. Jennings, J. Terpenny, R. Gao, and S. Kumara. A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering*, 139, 04 2017.

[41] W. Wu, Y. Zheng, K. Chen, X. Wang, and N. Cao. A visual analytics approach for equipment condition monitoring in smart factories of process industry. 04 2018. doi: 10.1109/PacificVis.2018.00026

[42] J. Xia, W. Chen, Y. Hou, W. Hu, X. Huang, and D. Ebertk. Dimscanner: A relation-based visual exploration approach towards data dimension inspection. pp. 81–90, 10 2016. doi: 10.1109/VAST.2016.7883514

[43] J. Xia, F. Ye, W. Chen, Y. Wang, W. Chen, Y. Ma, and A. Tung. Ldsscanner: Exploratory analysis of low-dimensional structures in high-

dimensional datasets. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 08 2017. doi: 10.1109/TVCG.2017.2744098
[44] F. Zhou, X. Lin, C. Liu, Z. Ying, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22, 11 2018. doi: 10.1007/s12650-018-0530-2

6.2 P2: RfX: A Design Study for the Interactive Exploration of a Random Forest to Enhance Testing Procedures for Electrical Engines

Full reference: *J. Eirich, M. Münch, D. Jäckle, M. Sedlmair, J. Bonart, and T. Schreck. Rfx: A design study for the interactive exploration of a random forest to enhance testing procedures for electrical engines. Computer Graphics Forum, pp. 1–14, 03 2022*

DOI: 10.1111/cgf.14452

COMPUTER GRAPHICS *forum*
Volume 0 (2022), number 0 pp. 1–14

RfX: A Design Study for the Interactive Exploration of a Random Forest to Enhance Testing Procedures for Electrical Engines

J. Eirich,[1,2] M. Münch,[2,3] D. Jäckle,[2] M. Sedlmair,[4] J. Bonart[2,5] and T. Schreck[6]

[1]University of Bamberg, Bamberg, Germany
[2]BMW Group, Munich, Germany
[3]Universtiy of Ulm, Ulm, Germany
[4]University of Stuttgart, Stuttgart, Germany
[5]Fraunhofer IWU, Dresden, Germany
[6]Graz University of Technology, Graz, Austria

Abstract
Random Forests (RFs) are a machine learning (ML) technique widely used across industries. The interpretation of a given RF usually relies on the analysis of statistical values and is often only possible for data analytics experts. To make RFs accessible to experts with no data analytics background, we present RfX, a Visual Analytics (VA) system for the analysis of a RF's decision-making process. RfX allows to interactively analyse the properties of a forest and to explore and compare multiple trees in a RF. Thus, its users can identify relationships within a RF's feature subspace and detect hidden patterns in the model's underlying data. We contribute a design study in collaboration with an automotive company. A formative evaluation of RFX was carried out with two domain experts and a summative evaluation in the form of a field study with five domain experts. In this context, new hidden patterns such as increased eccentricities in an engine's rotor by observing secondary excitations of its bearings were detected using analyses made with RfX. Rules derived from analyses with the system led to a change in the company's testing procedures for electrical engines, which resulted in 80% reduced testing time for over 30% of all components.

Keywords: human–computer interfaces, interaction, visual analytics, visualization

CCS Concepts: • Human-centred computing → Visual analytics; Systems and tools for interaction design; Graphical user interfaces

1. Introduction

In this paper, we contribute a design study on the use of *Random Forest* (RF) visualization for analysing data from test stations in a manufacturing process of the automotive industry. We address the problem of detecting faulty electrical vehicle parts that do not meet the quality requirements and thus need to be excluded from the manufacturing process. With the fast up-ramping of new assembly lines for electric vehicles across the globe [HKB*19], this problem has gained much attention across industries recently. To detect faulty parts, engineers typically rely on recording and analysing signal measurements, as widely used in many other engineering domains [SMF*20, EJS*20].

Traditionally, engineers analyse one measurement at a time and compare it to a ground truth baseline measurement. However, with the advent of new technologies, this one-at-a-time approach does not scale anymore. For testing modern electrical engines, for instance, engineers in our design study need to analyse more than 20 partially dependent signals within a single test procedure. Apart from the large burden of increasing manual work, it is also not possible to understand more complex patterns in the measurement data. At the moment, interaction effects between multiple measurements are completely left out. Without detecting issues stemming from such more holistic problems, however, faulty parts might not be detected (false negatives) or a high number of correct parts might be incorrectly labelled as faulty (false positives).

An apparent solution to deal with such situations is to leverage machine learning (ML)-based classification approaches. With these models, one can automatically explore many different combinations of measurements and see which combinations led to errors. This approach works specifically well if the objectives that define an error are well-defined and when the user is knowledgeable about ML [SMM12]. Both assumptions do neither apply to the users of our design study nor other design studies outside our application domain, for example in clinical research [BBJ*17] or music classification [RAZ*18]. Past examples in our application domain [SIB*11] have already shown that error detection is inherently ill-defined and needs the tacit knowledge of domain experts to be spotted. Similarly, this target audience usually comprises a high degree of domain knowledge, while dedicated expertise in data analysis might not be available [ACKK14, Via13].

Our design study focuses on a specific ML technique, RF [Bre01], which are among the more popular ML methods [HH15]. RF is an ensemble learning method that usually includes hundreds of independent *Decision Trees* (DTs). RFs are relatively easy to implement, less prone to overfitting than similar approaches such as DTs, and well-suited to detect dependencies in their feature subspace. They also serve well to find combinations of features from measurements that improve part testing, making them a good choice for our target domain [EJS*20, PBLG19]. However, without proper visual interfaces, RFs can be hard to interpret for users, specifically, for those who have little or no ML background. Efforts were made that sought to make RFs more interpretable through visualization [HH15, HWWH19, BBM*15, YXZC12]. As none of the existing approaches have been tested within a real-world scenario yet, our design study is meant to add ecological validity to this design space.

As a result, we present *RfX* (*Random Forest Explorer*), a visual analytics (VA) system that resulted from a design study project [SMM12] with automotive engineers at BMW. *RfX* aids domain experts with no data advanced analytics background, such as ML model building, in the interactive analysis of RF models. The system allows identifying DTs that contain the most promising measurement combinations for new test procedures, while users are able to interactively explore, compare and enhance DTs in light of their domain knowledge. In this context, we aim to find a solution to the following problem *'How can optimal rules be derived from an RF by domain experts with no background in advanced data analytics?'* In summary, our contributions are as follows: (1) the problem characterization and abstraction of our use case; (2) *RfX*, a VA system for the interactive visualization of an RFs decision-making process and (3) a field evaluation of our approach with five automotive engineers.

2. Related Work

In this section, we present approaches to visualize RFs and DTs, and present completed design studies in the automotive industry.

2.1. Random forest visualization

In a recent study, Hänsch *et al.* [HWWH19] distinguish between four groups of classical RF visualizations. *Abstract visualizations* show the model itself. *Result-driven* visualizations focus on the result of an RF [HH15] through the provision of graphical interpretations of the systems classifications. *Data-driven* visualizations map model outputs to their input data [SJC08]. *Parameter-driven* visualizations show internal parameters, such as specific features or split criteria [HH15]. Even though various works have aimed at visualizing RFs within or by combining the above-mentioned categories [SJC08, KvdWVW01, BBM*15], barely any tool allows the interactive exploration of several RF's properties, as well as the effects of parameter changes on the model's structure and output [HWWH19]. Exceptions are provided by Hänsch and Hellwich [HH15] and Yanh *et al.* [YXZC12], using a 3D representation to create a botanically inspired forest. However, one drawback of 3D visualizations are views that contain occlusion and perspective distortion effects [CCF97]. To tackle this issue, Hänsch *et al.* [HWWH19] render individual trees in a 2D space in a follow-up research paper.

The above-mentioned studies can result in very complex visualizations, especially when many DTs are used. Thus, other researchers suggest identifying most representative regions within an RF [WL19, BB17] and group similar DTs. This is important, because this means that it is not necessary to analyse ever tree in detail. Instead, only trees that represent a group of trees should be considered. Bakirli *et al.* [BB17] summarize recent studies on estimating tree similarities into the two categories: *semantic* and *structural* similarities. *Semantic similarity* [NKT08, ZJ12] measures internal tree properties, such as common feature subspaces between DTs for particular decision classes. In turn, *structural similarity* (also referred to as *syntactic similarity* [WL19, MS04]) compares structural tree attributes such as nodes, branches or number of leaves [Per11, Per13]. In this regard, Bremm *et al.* [BLH*11] develop a similarity score by combining leaf-, element- and edge-based measures. Dogra and Kobti[DK13] provide a method for finding similarities between knowledge represented by different DTs. Miglio and Soffritti [MS04] created an algorithm for tree similarity that combines tree structure information with agreement percentages on the testing set. Weinberg and Last [WL19] select the most representative DT from ensembles of DTs in big data environments. All of the above-mentioned approaches do either focus on the representation of the entire forest or the grouping of multiple DTs. While the former often results in complicated visualizations when models are very complex, the latter does not consider any kind of visual support at all, which makes an interpretation for domain experts with no analytics background cumbersome or even impossible.

2.2. Decision tree visualization

As an integral part of an RF's visualization, DTs must be visualized in a thoroughly adequate way. In this regard, von Landesberger *et al.* [LKS*11] give an overview of techniques that are appropriate for displaying DTs (*Node-link diagrams, treemaps* and *icicle plots). Node-link* techniques are probably the most well-known and often used tree visualizations [WFH*01, XHC*07, EW11]. In these approaches, nodes are represented as glyphs and relationships as links from parent to child. These visualizations can lead to scalability problems, especially when applied to large DTs. Thus, they are often combined with other visualization techniques. For instance, Bremm *et al.* [BLH*11] combine node-link diagrams with distance matrices, Behrisch *et al.* [BKSS15] with scatterplot views and Elzen *et al.* [EW11] with streamgraphs. As introduced by Schneiderman [Shn92], *Treemaps* are rectangular shapes that recursively subdivide rectangular spaces according to an underlying hierarchy. Even though they make good use of the available space, their nodes often overlap, which can lead to difficult distinctions of hierarchical structures [LKS*11].

By contrast, *icicle plots* place child nodes next to their parent and have been widely used to visualize DTs. For example, Ankerst *et al.* [AEK00] and Liu and Salvendy [LS07] use icicle plots to visualize DTs with continuous data. They also show class distributions by colouring the bars according to class values. Even though icicle plots do not make use of the entire available space, their advantage is that the parent nodes do not overlap with their children. Von Landesberger *et al.* [LKS*11] point out that the combination of *node-link diagrams* with *icicle plots* or *treemaps* increases the interpretability of a DT and thus allows for flexible analysis. All of the above-mentioned studies describe DTs outside the context of an RF. In addition, none of them has been validated within a large industrial setting or in our application domain.

2.3. Design studies in the automotive sector

Design studies in the automotive sector are mostly carried out for engineering design and anomaly detection. In the case of engineering design, efforts have been made to visualize in-car communication networks [SIB*11, SFMB12] or the exploration of multi-criteria alternatives for rotor designs [CMMK20]. Recent studies have been carried out making use of anomaly detection to detect error-prone produced parts in test stations of large-scale manufacturing processes [SMF*20, EJS*20]. While some of the mentioned studies acknowledge the need for interpretability of applied ML models [EJS*20], considerably less work has been dedicated to the development of systems that enable the analysis of such models. Grounded on previous findings, we did build such a system, that visualizes the decision-making process of an RF classifier.

3. Methodology

In this study, we primarily followed the nine-stage framework for conducting design studies of Sedlmair *et al.* [SMM12]. Taking the system design into consideration, we additionally used the *Design Triangle* by Miksch and Aigner [MA14] and the *Nested Model* for visualization design and validation by Munzner [Mun10]. As means of gaining access to experts with domain knowledge and to better comprehend the problem, we developed the system in close cooperation with two domain experts at BMW, who focus on the development of testing procedures for *electrical engines*. Both have a background in mechanical engineering and reported not being familiar with data advanced analytics methods, such as the training of an RF. We carried out a formative evaluation, in which we discussed iterative prototypes with the experts, who guided us during the development of *RfX* and provided us with feedback about and relevant tasks. A summative evaluation of the system was carried out with five domain experts from BMW. Methodological details for this downstream evaluation are provided in Section 8.

4. Problem Characterization

We now analyse the domain problem, the underlying data characteristics and relevant tasks.

4.1. Domain problem

The main problem we focus on is to support engineers in detecting measurement errors in part testing. To ensure high-quality standards, each part is tested through test stations in serial manufacturing processes, before it is allowed to be assembled into a car. Domain experts currently develop test procedures for automated testing of produced parts during different product development stages. In the early development stages, tests are developed in lab experiments by evaluating sensors for measuring product behaviour in specific circumstances. An example is torque behaviour during speed ramps of an engine to test its durability. Recorded data mainly comprises time series data that are analysed manually. Here, a manual analysis is feasible since part numbers are very low.

When a test is capable of detecting a possible error, for example a faulty gear of an engine, the test procedure is finished and deployed to its respective test station. New errors may appear that were not anticipated during product development phases, however, especially in the early stages of a test stations installation. In this stage, it is hard to adjust the already assembled test stations (e.g. installing new sensors is very costly), resulting in the need to use existing sensor equipment to detect new errors. To address this problem, engineers seek to detect new errors through analysing combinations of existing signal measurements. Currently, this is done manually by sequentially checking different combinations of signals, based on hypotheses that the engineers derive from their domain knowledge. Due to the high number of produced parts in this stage, however, these analyses result in lengthy and tedious processes. For example, one domain expert is responsible for the development of an end of line test station for *electrical engines*. However, this station is testing hundreds of engines per day, recording 1024 measurements where each measurement is capable to detect an error.

The main idea behind our approach is to leverage ML to detect more complex errors. ML naturally lends itself to this problem, since common ML techniques, such as RFs, are well-suited to map multiple features: from measurements to detected errors. In the course of discussions with our two lead users, however, it became clear that they have little to no ML expertise at the moment. As one user commented when we showed him classifier results *'Can you please explain to me why the model predicted this kind of error and say exactly which were the most important factors for this decision'.* While training the users in ML might be a natural solution, this is not always possible or realistic; we explore how far visual interfaces can go towards making DTs more accessible to non-ML experts. The goal lies in enhancing existing test procedures for the detection of produced parts that do not meet quality requirements and go beyond simple baseline tests. Here, we aim to enable domain experts to derive rules from an RF that are capable of enriching current test procedures. A simplified rule can be: '*If measurement A is greater than a given threshold X, then evaluate measurement B as a next step'.* Here, an analysis of an RF does not necessarily need to result in an optimized classifier that fits well to predict certain errors, but '*good enough*' to enhance existing test procedures.

4.2. Data abstraction and sample datasets

As input data for our visualization we use a binary RF classifier, according to Breiman [Bre01]. This model is appropriate because it is well-suited for the handling of high dimensional linear dependent data and is comparatively less complex than other ML approaches. Each RF consists of multiple DTs, whereby their number can be specified manually when setting the hyperparameters for the forest. Each DT consists of nodes and edges, where in our case, each node has always two children. An internal node consists of a feature from a measurement and a split threshold. Generally, it is hard to find out the exact number of DTs of each RF or the maximum depth of a DT in an RF since they depend on the properties of use case-specific data. We thus adapted these parameters in collaboration with our domain experts for the data sets we used in this design study. In addition, our classifier was constructed with the following hyperparameters:

First, we apply *bootstrapping* to overcome the common issue of the *bias-variance-tradeoff* [MRS08], often resulting in misclassifications. The maximum number of features in a tree are limited to $\sqrt{\#features}$ of all available features. As a split criteria, we use the *Gini importance* [LWSG13]. These hyperparameters are important for the abstraction of an RF since they give a general structure of tree behaviour. When using this hyperparameter set, each DT in the RF chooses its features randomly, splits the classes according to the *Gini importance* and at minimum five samples in a leaf. For an initial prototype of our system, we trained an RF with the *wine data set*, a widely used data set for ML use cases [ZF07, TD04]. We used this prototype for first discussions with our two lead users to gather first feedback on how to improve the visualization design. After our prototype was completed, we applied it to two datasets from our manufacturing setting. For each dataset, we analysed sensor data from two different test stations. Each sensor records acoustic measurements and comprises of multivariate time-series data. Detailed information about the feature extraction process can be found in Eirich *et al.* [EJS*20]. The two datasets are briefly described in detail.

1. The first dataset contained an RF trained with 219 features from electrical engines. An electrical engine is the final engine, which is assembled into a car and contains an electrical machine, a gearbox and an inverter. Features were extracted acoustic measurements of 900 randomly selected electrical engines produced between February 2020 and July 2020. Measurements were recorded at an end of line test station and served to analyse eccentricities of the rotor of an electrical engine. As an error class, 400 engines contained high eccentricities. This dataset was used in the evaluation with our domain experts to evaluate the usability of our system and the validity of our approach.

2. The second dataset contained an RF trained with 66 features from electrical machines. An electrical machine contains a rotor and a stator and is assembled into an electrical engine. Features were extracted from acoustic measurements of 200 randomly selected electrical machines produced between September 2020 and December 2020. Measurements were recorded at an inline test station and served to analyse the behaviour of the bearing inside an engine. As an error class, 100 electrical machines contained anomalies in the A-Bearing. This dataset was used above all to add further ecological validity to our approach with a second use case. We will also use this dataset as a running example in this paper, where 60 test samples from the initial dataset are visualized in Section 6.

4.3. Tasks

We began with the task characterization by observing our domain experts first using the prototypes and later the finished system. We focus on tasks to derive rules from promising feature combinations of multiple DTs. To derive relevant tasks, we used the taxonomies of Brehmer and Munzner [BM13] and Sedlmair *et al.* [SHB*14]. The former is appropriate because it forms the bridge between high- and low-level tasks, which is particularly helpful to embed them into the daily routines of the domain experts. The latter is well-suited because it particularly considers tasks that are relevant for ensemble modelling approaches such as RFs. Following the experts, we identified the following tasks **(T)**:

T1 *Partition RF*: To derive a rule from an RF, each analysis starts with the partitioning to find one single cluster of similar DTs or to find and compare multiple clusters of similar DTs within the RF. It includes an inspection of clustered DTs and their properties to answer questions like *'Are there visible groups of similar DTs in the RF visualization?'*. The task is also about gaining a first overview of the data space to answer questions such as *'How many groups of DTs exist?'* or *'How many DTs are within each group?'*

T2 *Identify individual representative DT*: Next, the decision to select a DT that is representative for a cluster within the RF is made. This task includes two sub-tasks. First, browsing (**T2.1**) through alternative similar DTs within a selected cluster and inspecting their high-level attributes, such as accuracy or number of nodes, allows identifying a subset of trees of special interest. For example, the accuracy of a DT can lead to the decision to select or skip a DT. This is followed by a comparison phase **T2.2**, where users primarily judge the superiority of one DT over other identified trees. Of course, the execution of (**T2**) involves going back and forth between the sub-tasks.

T3 *Explore individual DT*: When an interesting DT is selected, its low-level attributes, such as split thresholds or feature names are analysed in detail. For example, the name of a node and its child together with the split threshold can be directly mapped to a rule expressed as the statement: *'If measurement A > X, then evaluate measurement B'.* The resulting rule can be verified by reviewing the distribution of the classes in each node. This task is related to *Uncertainty* [SHB*14] since it involves the evaluation of the reliability of the model output mapped to the real data instances.

T4 *Optimize DT*: The split thresholds can be adapted in the light of previous domain knowledge. Thus, a suggested rule from a DT can be iteratively refined until all split thresholds are properly adapted. We are aware of the fact that individually optimized DTs might result in overfitted models. Thus, each rule that contains adapted thresholds has to be tested with new training data.

5. Data Aggregation for the Visualization

Our approach is closely related to *ensemble pruning*[TPV09, GRF00, DSM16, ZZY17]. In this method, an ensemble size of classifiers is reduced to increase model efficiency and predictive performance [BBSH14]. While efficiency and performance are undeniably important aspects of model creation, the interactive exploration of individual ensembles in the context of an RF remains the focus of this work. Our approach does not aim to exclude any DT from an RF. Instead, we provide an approach for clustering similar DTs within an RF, allowing the user to explore multiple DTs. Many previous studies showed that a combination of semantic and structural similarities resulted in an efficient grouping of similar DTs [WL17, WL19, BB17, MS04]. We build on this existing knowledge and combine semantic and structural tree similarity scores. The commensurate score of the DTs (T_i and T_j) is computed via a convex combination in Equation (1). To group similar trees, we cluster the resulting matrix. As input data for *RfX*, the matrix with the highest cluster accuracies, that lead to an optimal parameter λ^* in Equation (1), is selected, as outlined in Section 5.3. Instead of other metrics mentioned in the literature (Section 2.1), our metric specifically considers the context of an RF. This is important since ensemble-based DTs differ from DTs in an RF. For example, DTs from an ensemble are generally deeper and thus more susceptible to overfitting.

$$Sim\big(T_i, T_j\big) := (1-\lambda)\, Sim_{\text{stru}}\big(T_i, T_j\big) + \lambda\, Sim_{\text{sem}}\big(T_i, T_j\big), \tag{1}$$

5.1. Semantic dissimilarity

As one possible score to compute semantic similarity, Weinberg and Last [WL19] measure how often different DTs agree on the test data and assign a similarity score based on the result. Instead of other semantic similarity measurements to compare DTs, such as the *Jaccard Index* [FI18], Weinberg and Last [WL17, WL19] showed in various studies that agreement of DTs is a valid similarity score. Hence, we consider their approach as appropriate for our use case. When evaluating this approach, however, our users reported that they were confused to find that high scores represent DTs that are close to each other. Thus, we inverted the similarity score of Weinberg and Last to a *dissimilarity score* so that the disagreements are counted in Equation (2), where a low score is assigned to a pair of DTs if they often agree and therefore close to each other. For Equation (2), D_{Test} is the set of test samples and $\mathbb{1}_{\{T_i(\mathbf{x})\}}(\cdot)$ is the indicator function of the set $\{T_i(\mathbf{x})\}$.

$$\sum_{\mathbf{x}\in D_{\text{Test}}} \big(1 - \mathbb{1}_{\{T_i(\mathbf{x})\}}\big(T_j(\mathbf{x})\big)\big) \tag{2}$$

To compare both structural and semantic scores, we include a scaling interval. Thus, dissimilarities are scaled to the interval [0,1]. To do this, the semantic score of each pair of trees is divided by the maximum score of two trees within the RF. These values are stored in a dissimilarity matrix $Sim_{\text{sem}} := (Sim_{\text{sem}}^{(i,j)})_{i,j=1}^N$, with

$$\begin{aligned} Sim_{\text{sem}}^{(i,j)} &:= Sim_{\text{sem}}\big(T_i, T_j\big) \\ &:= \frac{\sum_{\mathbf{x}\in D_{\text{Test}}} \big(1 - \mathbb{1}_{\{T_i(\mathbf{x})\}}\big(T_j(\mathbf{x})\big)\big)}{\max_{p,q\in\{1,2,\ldots,N\}} \sum_{x\in D_{\text{Test}}} \Big(1 - \mathbb{1}_{\{T_p(\mathbf{x})\}}\big(T_q(\mathbf{x})\big)\Big)}. \end{aligned} \tag{3}$$

5.2. Structural dissimilarity

Furthermore, we build on the work of Bakirli *et al.* [BB17] to group paths in a tree by their output labels and transform these groups into respective sequences. To be consistent with Equation (2), we refer to this score as structural dissimilarity instead of similarity. We use the Levenshtein distance [Lev66] to measure the dissimilarity between each sequence, which is a popular character-based metric and measures the minimum number of single-character edits between two words. This fits in well with our approach because branches of DTs can be reduced to bitstrings, which is a perfect basis for interpreting the differences as dissimilarities.

For each tree T_i, we extract its *branches* and group them according to the *output-labels*. This assigns each tree T_i a representation as two *tree sequences* (TS)s. Each tree sequence $TS_{i,l}$ consists of $m_{i,l} \in \mathbb{N}$, the number of paths in the DT T_i that end with the label l, *branch sequences* (*BS*)s. The branch sequences are representations of the paths within the tree and written as $(s_k + 2)$-tuples, where $s_k \in \mathbb{N}$ denotes the length of the kth branch, excluding root and leaf-nodes. The root of tree T_i is denoted as r_i and $f_{i,k,n}$ is the feature of the splitting node n of the kth branch of the respective tree sequence $TS_{i,l}$.

The root and the features are coded as elements of the alphabet, which we extend by *AA*, *AB*, *AC* and so on if needed, corresponding to their feature names in the training data. A *BS* terminates with the corresponding label $l \in \{0, 1\}$ that gives the output of this branch in the DT. We order the *BS*s of the same label within a DT from the left to the right. For example, a branch can have the sequence A-C-E-1, where 1 represents the class and A-C-E the features splitting the nodes. Next, we concatenate all of the branches for every tree with the same label along our ordering from the left to the right. This gives us two sequences, each of which is equivalent to its corresponding TS. These new sequences have the form

$$TS_{i,l} := \Big(r_i, f_{i,1,1}, \ldots, l, r_i, f_{i,m_{i,j},1}, \ldots, l, f_{i,m_{i,j},s_{m_{i,j}}}, l\Big). \tag{4}$$

Analogously to Section 5.1, two trees have a small dissimilarity score, if they are similar and a high score if they are not similar. We also scale the structural dissimilarity matrix, by maximum scaling, to the interval [0, 1]. This implies that the final dissimilarity matrix (see Equation 1), as a convex combination of the two specialized matrices, is also scaled in the interval [0, 1]. Since the semantic and structural matrices are symmetric, the resulting dissimilarity matrix as input for our visualization is also symmetric.

5.3. Clustering and two-dimensional representation

With both dissimilarity matrices available, we cluster similar DTs. Since these matrices depend on the parameter λ, the content of this section is always dependent on this parameter. However, we suppress the dependency in the following notations. The optimal parameter λ^* for Equation (1) is computed at the end of this section. We choose a clustering approach, where the number of clusters is automatically determined. A widely used method for this is a combination of *k-means* [Mac67] for clustering and an evaluation of the optimal number of clusters by *Silhouette Coefficients* [KLKR90]. This is an appropriate choice over other clustering approaches, such as clustering with DBSCAN, which needs an extensive analysis of the parameters to achieve a good cluster structure. We name the resulting clusters $C_1, \ldots, C_k$. Each DT T_i is represented as a vector

$t_i \in \mathbb{R}^N$, where t_i is the ith row of the dissimilarity matrix and N is the number of DTs in the RF. With that, each DT in the RF is assigned a cluster-label. Next, we project our dissimilarity matrix into a 2D space via *Multidimensional Scaling (MDS)* [Pea01]. Since the accuracy (acc_i) of a tree is an important high-level attribute, we want to take this measurement into account when analysing the clustering. Thus, we expand the 2D representation of each t_i by its accuracy to be $\tilde{t}_i := (x_i, y_i, acc_i)^T$ and compute the arithmetic means $\bar{t}_u$ of every cluster C_u such that

$$\bar{t}_u := (\bar{x}_u, \bar{y}_u, \overline{acc}_u)^T := \frac{1}{|C_u|} \sum_{\tilde{t} \in C_u} \tilde{t}, \tag{5}$$

with $|C_u|$ being the cardinality of cluster C_u. The two-dimensional visual representations of the cluster centres consider the accuracy by weighting the vectors $(\bar{x}_u, \bar{y}_u)^T$ by $\overline{acc}_u$. This gives us the centre c_u of cluster C_u in Equation (6).

$$c_u := \left(c_u^x, c_u^y\right)^T := (\bar{x}_u \cdot \overline{acc}_u, \bar{y}_u \cdot \overline{acc}_u)^T \tag{6}$$

With the 2D projections of the DTs, we then find for each cluster C_u the DT that is closest to the theoretical centre c_u as

$$t_u^* := \arg\min_{\tilde{t} \in C_u} \sqrt{\left(x - c_u^x\right)^2 + \left(y - c_u^y\right)^2}. \tag{7}$$

In order to obtain more detailed information, we repeat these clustering steps on every individual cluster to get sub-clusters with their representative trees. Here, the number of sub-clusters of each C_u is also automatically determined by evaluating Silhouette Coefficients. Thus, the RF is divided into clusters and each cluster into sub-clusters. When we define S as a set of trees, then its classification output $l_{S,\mathbf{x}} \in \{0, 1\}$ for a test sample $\mathbf{x}$ is computed in Equation (8). With this, the *accuracy* acc_C of a cluster C is defined in Equation (9), where $\mathbb{1}_{\{l_\mathbf{x}\}}(\cdot)$ is the indicator function of the set that consists only of the sample $\mathbf{x}$'s label $l_\mathbf{x}$. Thus, each cluster can be seen as little RF within the RF.

$$l_{S,\mathbf{x}} := \arg\max_{l \in \{0,1\}} |\{T \in S | T(\mathbf{x}) = l\}|. \tag{8}$$

$$acc_C := \frac{1}{|D_{Test}|} \sum_{\mathbf{x} \in D_{Test}} \mathbb{1}_{\{l_\mathbf{x}\}}\left(l_{C,\mathbf{x}}\right), \tag{9}$$

We use Equation 9 to determine the parameter-value λ^* whose resulting dissimilarity matrix is input to *RfX*. This value is chosen among the values $\{0, 0.1, \ldots, 1\}$ to have the highest *mean-accuracy* of the clusters that is

$$\lambda^* := \arg\max_{\lambda \in \{0, 0.1, \ldots, 1\}} \frac{\sum_{C \subset RF, C\ cluster} acc_C}{|\{C | C \subset RF, C\ cluster\}|}. \tag{10}$$

6. *RfX's* Visualization Components

Our visualization design is shown in Figure 1 and comprises six components labelled from (a) to (f). In general, the workflow of the system is closely related to a *multiattribute choice* as defined by Dimara *et al.* [KLKR90]. In our case, it can be described as finding one or multiple good-performing DTs among a finite number of other DTs and comparing them with each other. This is achieved by inspecting several high- and low-level tree attributes (Views (a)–(d)) of each selected DT. When a DT is considered relevant for the user, one or multiple rules can be derived, verified, and enhanced by the user (Views (e) and (f)_). For the entire demonstration of our visualization, we use the 60 test samples from the second test dataset from Section 4.2 as a running example.

6.1. Random forest view

The scatterplot in (a) helps to partition an RF (**T1**) by providing a big-picture of all available DTs and represents an entry point into an analysis. We use a planar projection, which is a common analysis start to get an initial overview over high-level relationships [SZS*16, JSM*17]. Each DT is represented by a single dot, while the size is increased for highly accurate trees and decreased accordingly. The cluster id and the accuracy are displayed in additional rectangles for each cluster. By hovering over a tree, its low-level attributes are displayed (e.g. accuracy).

6.2. Icicle plot and table view

To facilitate the identification of individual representative trees within clusters (**T2**), we encode them as small multiples [Tuf90]. Representative trees are identified as outlined in Section 5.3 and shown as icicle plot in (b). This view allows the inspection of individual high-level tree attributes (e.g. decision paths), where additional attributes can be inspected via hovering over each icicle plot (e.g. observations in classes 1 and 2). Each row is mapped to the parent cluster via rectangles with the same colours as the clusters in panel (a). Rows are sorted according to the accuracy of all clusters and columns to the accuracy of DTs within a cluster. The class distribution in each rectangle is shown on a divergent colour scale between red (Class 1) and blue (Class 2).

Some users preferred tables over icicle plots. Hence, we added table views for all clusters in (d) and all trees in a cluster in (c). As well as (b), the tables have the purpose to identify a relevant DT (**T2**). Tables are sorted according to their accuracies. The selection of a cluster in (d) and a DT in (c) results in their highlighting in (b) (see the background of clusters for a selected cluster and red dot for selected DT).

6.3. Decision tree view

Each tree is displayed as a node-link diagram in (e), these diagrams are widely used to visualize DTs [EW11, LKS*11, BLH*11]. As high-level tree attributes the tree id, its accuracy, the number of interior (non-leaf) nodes and its confusion matrix are displayed in the upper left of (e). Each cell of a confusion matrix is represented with two triangles, where the matrix rows represent the actual values and the columns the predicted values. To put focus on matrix cells that contain non-zero values, we increase the transparency of empty cells. This further facilitates the decision to select or skip a DT out of multiple DTs. Each node of the node-link diagram is shown as a confusion matrix, which represents—as another small

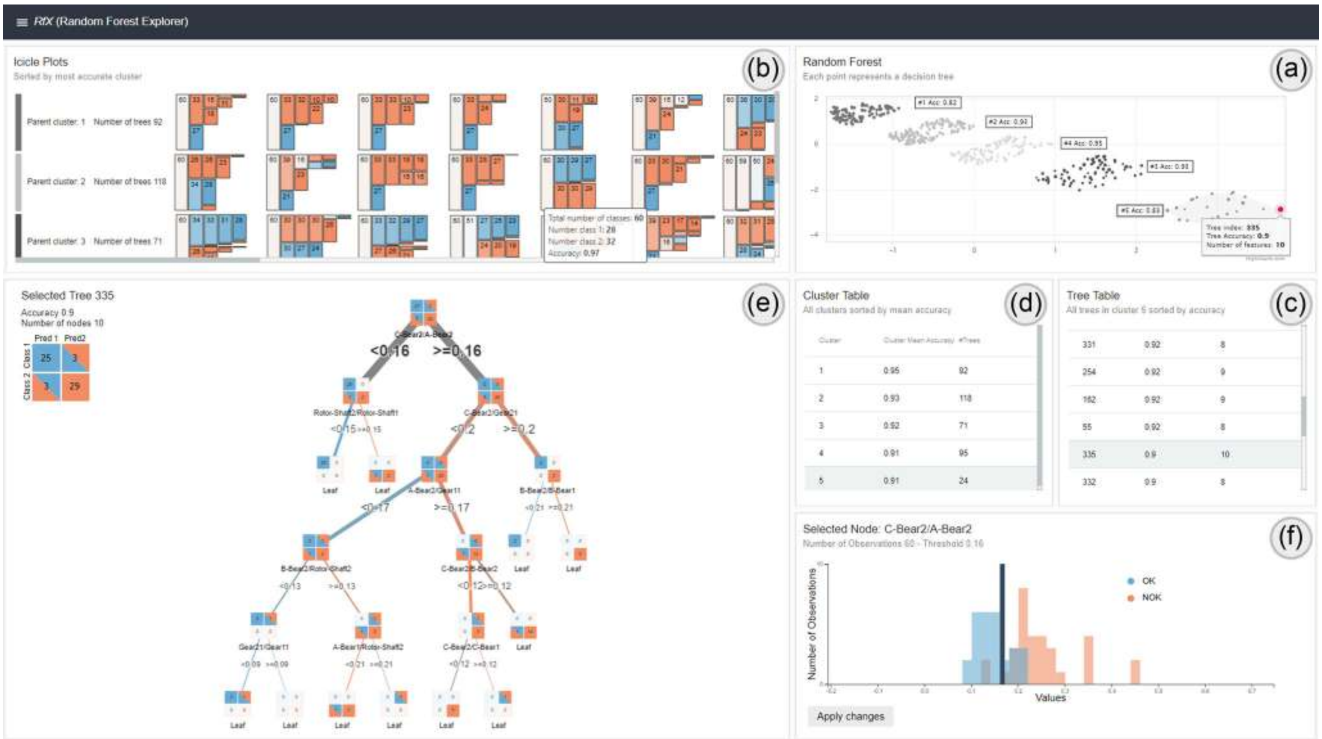


Figure 1: *Screenshot of* RfX *with 400 DTs from our second use case data-set as described in Section 4.2. (a) Visualizes the projection of our combined dissimilarity matrices. (b) Shows most representative tress of each cluster, represented as an icicle plot. Each cluster of similar trees is additionally visualized with a table representation, where (d) shows each cluster and (c) each tree within a cluster. A DT and its properties are shown as a node-link diagram in (e) and the class distribution of a feature as a histogram in (f).*

multiple [Tuf90]—the quality of the split criteria for the distribution inside the node. In addition to the icicle plots, edges are coloured in accordance with the same colour scheme, while split thresholds are displayed on the edges. The name of a feature is displayed under a node. Where previous views help to find appropriate DTs in the RF, the node-link diagram serves as the heart of the system to derive a rule from the DT (**T3**).

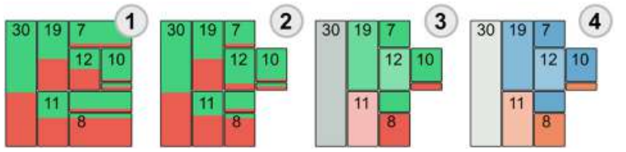


Figure 2: *Evolution of the icicle plots to represent a DT. In a first attempt in (1), we used a treemap, which we changed in (2) to an icicle plot with a qualitative colour scale. In (3), we used a divergent colour scale, which we changed in (4) to account for colour deficiency.*

6.4. Histogram view

In the histogram in (f), users can review the actual class distribution where classes are marked with the same colours as in (e) and (b). The number of bins depends on the length of the input data vector. Since in our case, the test data contains 60 samples, we also use 60 bins in the histogram. A histogram is displayed by selecting a node in the node-link diagram. If tree splits are not optimal, users can adapt thresholds individually by dragging the blue rectangle in (f) (**T4**). Thus, a promising rule discovered in (f) can be iteratively refined, while threshold changes are always mapped to all confusion matrices in the node-link diagram. Furthermore, adapting thresholds supports users in the interactive analysis of the decision-making process of the RF, which can result in new insights and knowledge for domain experts [SSS*14].

7. Iterative Design Process

The design of *RfX* was carried out in four main iterations in collaboration with the domain experts. After each iteration, we gathered feedback and adapted our visualization design.

Figure 2 shows the evolution of the icicle plot representation. First, we experimented with treemaps in (1). However, domain experts reported that it was hard to follow single branches of a tree since nodes overlapped on the horizontal axis. Therefore in (2), we used icicle plots, because these allow users to better identify and analyze single branches of DTs [LKS*11]. In a first attempt, we used a categorical colour scale to distinguish between two classes. Here,

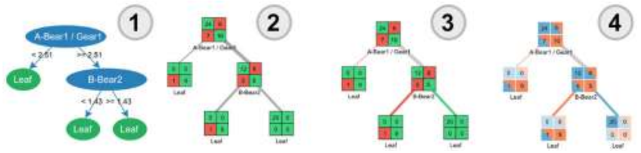


Figure 3: *Evolution of the node-link diagram to represent DTs. In (1), we use a simple node-link diagram. (2) Contains confusion matrices for each node and (3) colours for the edges to account for dominant classes in each split. In (4), we changed the colour to account for colour-blindness and represent each matrix cell as with two triangles (column for prediction and row for real value).*

domain experts reported that they are more interested in an efficient way to analyse the performance of a DT. They also found it difficult to evaluate the split quality of individual nodes in a DT. It was not always clear what the dominant class in a node was, especially in small rectangles with few samples. We thus used a divergent colour scale in (3). With such a colour scale, the dominant class in each node stands out. To ease the evaluation of DTs, we scaled the height of each icicle plot according to its accuracy. Since the colours in (3) did not account for colour deficiency, we adapted the colour scale to tackle this issue in (4).

Figure 3 shows the evolution of the tree views. In (1), we represented all nodes as ellipses. Although this is a common approach to represent DTs [SS18, LKS*11, BLH*11], experts found it hard to evaluate the distribution of samples inside each node. Thus, in (2), we introduced confusion matrices for each node where only false positives and negatives were coloured red. To show the distribution of class splits, we coloured the edges in (3) with the same divergent colour scale as we outlined above. To facilitate the analysis of each confusion matrix, we encoded each cell with two triangles as demonstrated in Section 6.3. Furthermore, we changed to a colour scale that accounts for colour-blindness. These changes allow experts to easily identify false positives, and negatives in a DT and analyse their distributions.

8. Evaluation

In this section, we introduce our evaluation methodology and demonstrate the usability of our system together with the application of found rules to a real-world scenario.

8.1. Methodology

Our evaluation aims at validating the usefulness of the proposed technical considerations (Section 5) and the resulting visualization (Section 6) in terms of effectiveness and problem-solving characteristics for domain experts within their daily routines. We present the results of an expert study, embodying qualitative coding of user feedback in combination with a quantitative usability scale.

Participants: We conducted the study with five domain experts from BMW, responsible for the development of testing procedures for the serial manufacturing of electrical vehicles. They were all male between 23 and 33 years old and had a mean working experience of 2 years in the problem domain. Every expert had a background in mechanical engineering, electrical engineering or physics and hence a higher education. Each of the five also reported being unfamiliar with advanced data analytics methods.

Data: For the evaluation, we used the first dataset as described in Section 4.2. The measurements of the data were recorded to detect engines with increased eccentricities of the rotor. The testing method is split into a general noise vibration part and a dedicated eccentricity measurement. To detect engines with high eccentricities within the first phase, labels from the second testing phase were used to train the model. Due to non-disclosure agreements with BMW, we cannot mention the absolute error rates or the cycle times for each testing procedure.

Task: The following task was given to each user: *'Derive a rule from a feature combination that you found in a DT'.* During the development with our lead users, we observed that tasks generally were carried out according to the workflow *'overview first—then details on demand'* as described in Section 6. Thereby, an example of an execution can be as follows: first, select a cluster either with the scatterplot view or from the cluster list that seems relevant. Next, select a tree from the cluster either via the icicle plots, the scatterplot, or the tree list. Then, explore the DT, by clicking on its nodes to review the class distributions. Finally, interactively change the thresholds if class splits are not optimal. The expected outcome is a rule from a combination of features and thresholds that can improve a current test procedure.

Procedure: The study was conducted in the form of a think-aloud study [SBS94] with one observer taking notes. Each session took on average 90 min and involved a detailed and prescribed walk-through of the system, open-ended questions [Sau20] about its usage and a usability questionnaire. Since the task of the domain experts and the visualization design are both relatively complex, we made sure that both the concept of our combined dissimilarity matrix and each view of *RfX* was understood well before starting the think-aloud study. All interviews were performed online, where each participant executed the same predefined task. The notes taken during the think-aloud study were analysed using a qualitative coding methodology [Cha06]. To quantitatively assess the usability of our system, we applied the *System Usability Scale (SUS)* [Sau20]. This scale is composed of ten statements rated on a Likert scale. The qualitative coding scheme and the quantitative (SUS) provide a comprehensive picture of our tool's deployment readiness level.

8.2. Findings from the think-aloud study

After coding and sorting the participants' comments and our observations, we were able to derive insights about the usability of the system. Four of five users first used the scatter plot to get an overview of the distribution of the clusters and the closeness of trees within each cluster. The remaining user reported that sufficient information was visualized in the list views. Next, users selected a cluster because it either contained an overall high accuracy (three users) or a comparable high accuracy but fewer DTs (two users). After each user selected a cluster, we identified two workflows that were then carried out. First, three users used the table view instead of the icicle plots to select trees with a low number of nodes and

Table 1: *Results of the System Usability Scale [Sau20] with five domain experts responsible for electrical engines.*

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
Expert 1	7.5	10	10	10	10	10	10	10	7.5	10	95
Expert 2	10	7.5	10	7.5	7.5	10	10	10	10	7.5	90
Expert 3	10	7.5	7.5	10	10	10	5	7.5	7.5	10	85
Expert 4	7.5	10	7.5	5	7.5	10	7.5	7.5	7.5	10	80
Expert 5	7.5	7.5	7.5	10	7.5	10	7.5	7.5	7.5	7.5	80
Avg.	8.5	8.5	8.5	8.5	8.5	10	8.0	8.5	8.0	9.0	86

Table 2: *Comparison of confusion matrices and model accuracies of a RF, a DT, and a derived ruleset. All modelling approaches were built with new sensor data from new electrical engines, which were produced after the ones we used in our study. 5% of the data contained an error.*

		RF pred.		DT pred.		Rule pred.	
		OK	NOK	OK	NOK	OK	NOK
Real value	Ok	0.93	0.07	0.91	0.09	0.91	0.13
	NOK	0.00	1.00	0.09	0.91	0.00	0.94
Accuracy		0.956		0.911		0.922	

a high accuracy and compared them to trees inside and outside the cluster. They then used the meta information of each tree, such as the confusion matrix to find trees with low false negative rates. The remaining two users used the icicle plots to select a tree. Here, the first user selected the top left DT because it was the most accurate one, while the second user discovered a rule that he thought was interesting in the left DT of the second cluster and thus decided to analyse this DT in detail. For both workflows, a tree was reported as relevant, if it either contained a new rule that had not been noticed by the users before or had a low number of false negatives in the confusion matrix. When asked why some users considered icicle plots and others the tree lists as relevant for their choice to select a DT, the users reported that they either generally preferred lists over visual abstractions or vice versa.

After a tree was identified, all users reviewed the names of each node to evaluate if their connection to other nodes made sense for them, where one expert noted *'I expected to find signals that I know, but in fact did find signals, which I did not expect at all. This is good because it makes me think of new relevant relationships within the data'*. Here, the icicle plots, confusion matrices for each node and colours and thickness of each edge were noted as especially helpful (*'It is good to see where the main path inside the tree lies to get a general overview of its behaviour and to evaluate how it separates the classes')*. After selecting relevant nodes, all experts revised the histogram and adapted split thresholds, when they thought that better splits in terms of DT accuracy could be achieved. This resulted in better classification results of the selected DT. All users reported that it was important to explore different types of DTs, which all represented different rule sets. Many of the newly discovered rules were interesting because they showed relationships between signals that users had not thought of previously and which helped them to better understand eccentricities of the rotor.

8.3. Findings from the system usability scale

Taking the quantitative results of the usability survey into account, our system provides good usability according to the adjective equivalent of the achieved SUS score [BKM09]. With a score of 86, our system is well above the average score of 68 [Sau20]. The individual scores are outlined in Table 1. *RfX* scores highest on low inconsistencies (Q6) and lowest on the quick learning of the system (Q7) and confidence in system usage (Q9). A possible explanation for the results from Q7 and Q9 is that interpreting ML models without prior knowledge is by no means a simple task. However, we are not able to draw a final conclusion on this observation, because of the low number of participants involved.

8.4. Anecdotal evidence

As mentioned in Section 8.1, the detection of *electrical engines* with high eccentricities is divided into two phases. The first phase measures general noise vibrations and the second detects dedicated eccentricities. During the think-aloud study, eight promising feature combinations from five DTs were detected by the domain experts with *RfX* and translated into rules. Each rule was tested on new sensor data from completely new electrical engines, which were produced after the ones we used in our study, where 5% contained an error in the rotor system. Due to non-disclosure agreements, we are not able to mention absolute error rates of engines. The low number of faulty engines is because the production quality at our industrial partner is very high and only very few engines are produced that have an error. After testing each rule on this new sensor data, we found that two rules from two different DTs described how increased eccentricities in the rotor system can be detected by observing secondary excitations of its bearings. When combined with the existing testing procedure, over 30% of all engines with high eccentricities were detectable during the first testing phase. Thus, for these engines, the second longer testing phase can be skipped in future, reducing the testing time for engines that do not meet quality criteria by 80%. Furthermore, we wanted to know if these results could also be achieved with a simpler yet intrinsically interpretable model. Thus, we used the same dataset to train a DT as an interpretable model and an RF as a more complex model and compared the results in Table 2. Here, the RF did achieve the best accuracy, the derived rule-set the second best and the DT the lowest. Thus, the derived rule-set outperforms an intrinsically interpretable DT but not a more complex RF.

9. Discussion

In today's highly automated manufacturing processes, the extraction of relevant and meaningful information from high-dimensional data remains a challenging problem [BKSS15, EBJ*21]. In this regard, the cooperation between human experts and ML techniques has often proved to be a promising solution by combining the strengths of both worlds [SMF*20, JFSK15]. We contribute to this challenge with a design study and present the system *RfX*.

RfX provides interactive visual means for identifying relevant DTs within an RF, thus enabling the user to discover relationships and detect hidden patterns. Since identifying relevant DTs is a difficult task for users who do not have an advanced analytics background, the set of visual components we have introduced follows a **dynamic analysis process**. Here, users can interact with our visualization in different ways to explore the decision-making process of an RF and to gain new knowledge. For instance, the user can choose between an *exploratory* and a *target-driven* analysis of the DTs, based on the visual component they choose as an entry point.

The planar projection provides the *exploratory* entry point into the analysis. An overview of this kind enables the user to choose a cluster with high accuracy (e.g. by inspecting the rectangles containing additional information in Figure 1(a)) and trees that are located close to each other. The icicle plots then allow the user to immediately determine decision paths in each tree and to compare them to similar trees within the same cluster.

By contrast, the cluster table supports an efficient, *target-driven* analysis. The cluster table is sorted by the accuracy of each cluster and lets the user select a DT cluster that performs relatively well but contains fewer DTs. The user then selects DTs from the tree table to identify a DT with the lowest possible number of false negatives while minimizing false positives. Here, a DT may contain many false positives but still be relevant, since it contains fewer false negatives. This is especially relevant in manufacturing contexts where further assembly of false negative results in overall faulty parts [EJS*20]. Despite the fact that both analyses have the goal of selecting the most relevant DT, they follow different workflows, in which both address a multivariate choice [DBD17].

The challenge of achieving an optimal trade-off between false positives and negatives is also relevant for the adaptation of split thresholds. While some might argue that this kind of task is best addressed completely automatically, we argue that human domain knowledge is a valuable resource [ACKK14, Via13] that serves well to optimize ML techniques. For example, users might prefer to adapt thresholds, which on the one hand produce a higher number of false positives, but on the other hand, provide a solid threshold that minimized false negatives. In contrast, other users might want to optimize thresholds that both result in minimal false positive and negative rates but are prone to overfitting. In this regard, we also acknowledge that thresholds could be adopted without manual intervention during training by including suitable loss and regularization techniques. Furthermore, we want to stress the fact that an important aspect of our approach is also about newly gained knowledge of domain experts from the interactive exploration of the decision-making process of the RF. New knowledge is difficult to measure [EJS*20], which is why we believe that providing an interactive means to understand complex ML models is important and we are confident that our approach addresses this issue. Even though the main goal of our approach is not to optimize an ML classifier the most efficient way, we showed that human domain knowledge can be an important aspect in better understanding an ML classifier. This is demonstrated by the fact that the derived rule-set from the domain experts outperformed a simple DT in terms of accuracy (see Table 2).

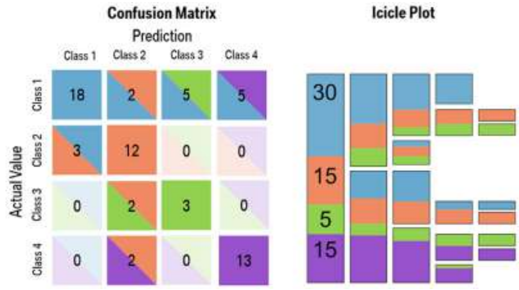


Figure 4: *Examples of a confusion matrix (left) and an icicle plot (right) for a multi-class prediction task using a qualitative colour scale.*

While similar systems of various different types so far have been proposed for exploring the decision-making process of ML classifiers, *RfX* differs from them in some important aspects. For example, Ming *et al*. [MQB19] provide the system (*RuleMatrix*) to derive a list of rules out of a neural network. However, their approach aims at approximating and analysing a single DT, whereas our approach groups similar DTs, leaving the possibility to explore, analyse and compare multiple DTs. Furthermore, *RuleMatrix* uses only flat representations (e.g. list views) and is not evaluated within a real-world context. In turn, *RfX* combines hierarchical (e.g. icicle plots) and flat (e.g. table views) representation and evaluates their interplay within a real-world scenario.

We acknowledge that our study has some limitations, which can be addressed in future research. First, our solution is a specific use case, designed and implemented for a specific problem in a specific company. Furthermore, our evaluation only involves five domain experts. This is a relatively common situation in design studies, where the presented visualizations often tackle very specific problems, which can be addressed by only a few experts [BSKR19, CMMK20, SIB*11, SFMB12, SMF*20]. A second limitation is that we only use a binary classifier. Even though many classification problems can be solved with binary classifiers a multi-class classifier leaves different challenges. To address this issue from a visualization perspective, the icicle plots could be enhanced with a qualitative colour scale to account for more than two classes. Using a qualitative colour scale for multi-class problems would also address the problem of imbalanced data. Here, classes with small sample sizes would still be visible in rectangles in the icicle plots. In the context of the representation of our confusion matrix, more classes could easily be added by also using a qualitative colour scale. An example of how a qualitative colour scale could address a multi-class problem for confusion matrices and icicle plots is provided in Figure 4.

Third, the experts also recommended one system improvement. The experts requested the inclusion of help buttons in the system (e.g. a detailed explanation of the icicle plots). Finally, neither the single concepts of our dissimilarity score (e.g. semantic dissimilarity) nor the visualization components (e.g. icicle plots) are novel as such. However, to the best of our knowledge, neither a combination of dissimilarity scores has been combined with well-established visualization components nor have similar approaches been tested within a real-world scenario. Thus, we regard this work as an exploratory step towards explainable ML in manufacturing settings and plan to further extend and validate the idea of exploring groups of similar DTs in an RF with interactive systems.

10. Conclusion and Future Work

In this paper, we present a design study for the system *RfX*, which allows the deriving of rules from an RF's decision-making process for users with no background in ML. To achieve this goal, we first build on existing work to develop a combined score for *semantic* and *structural* tree similarities that enable the grouping of similar DTs within an RF. Guided by previous work on the visualization of RFs [HH15, HWWH19] and by using state of the art visualization techniques to visualize DTs [LKS*11, EW11], users can partition an RF, identify relevant DTs within an RF, explore each tree individually and adapt tree thresholds in light of their domain knowledge. A validation through a field study with five domain experts from BMW backed our technical considerations and design choices. In addition, as a result of *RfX*'s use, a derived rule improved the overall manufacturing process to detect electrical engines with high eccentricities and resulted in a reduced testing time of 80% for over 30% of engines that did not meet quality criteria from the analysed organization.

Fuelled by the positive feedback from our users and the improvement of the analysed manufacturing processes, future research efforts will build on this foundation. One possible extension of *RfX* will be to add another component to extract and represent features from the data space. Our intention here, is to develop a method that automatically maps important features of a trained model to the existing visualization. One way to achieve this objective could for example, be to use the RF feature importance score or features that often appear in highly accurate trees.

Acknowledgements

We would like to thank all study participants at BMW for their great support while developing *RfX*. This project was founded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2075-390740016

Open access funding enabled and organized by Projekt DEAL.

References

[ACKK14] Amershi S., Cakmak M., Knox W. B., Kulesza T.: Power to the people: The role of humans in interactive machine learning. *AI Magazine 35*, 4 (Dec. 2014), 105–120.

[AEK00] Ankerst M., Ester M., Kriegel H. p.: Towards an effective cooperation of the computer and the user for classification. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Apr. 2000).

[BB17] Bakirli G., Birant D.: Dtreesim: A new approach to compute decision tree similarity using re-mining. *Turkish Journal of Electrical Engineering & Computer Sciences 25* (Jan. 2017), 108–125.

[BBJ*17] Bannach A., Bernard J., Jung F., Kohlhammer J., May T., Scheckenbach K., Wesarg S.: Visual analytics for radiomics: Combining medical imaging with patient data for clinical research. In *Proceedings of the 2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)* (2017), pp. 84–91.

[BBM*15] Beck F., Burch M., Munz T., Di Silvestro L., Weiskopf D.: Generalized pythagoras trees: A fractal approach to hierarchy visualization. In *Proceedings of the VISIGRAPP - Computer Vision, Imaging and Computer Graphics - Theory and Applications* (Jan. 2015), pp. 115–135.

[BBSH14] Bhatnagar V., Bhardwaj M., Sharma S., Haroon S.: Accuracy-diversity based pruning of classifier ensembles. *Progress in Artificial Intelligence 2*, 2–3 (2014), 97–111.

[BKM09] Bangor A., Kortum P., Miller J.: Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies 4* (Apr. 2009), 114–123.

[BKSS15] Behrisch M., Korkmaz F., Shao L., Schreck T.: Feedback-driven interactive exploration of large multidimensional data supported by visual classifier. In *Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology (VAST)* (Feb. 2015), pp. 43–52.

[BLH*11] Bremm S., Landesberger T., Hess M., Schreck T., Weil P., Hamacher K.: Interactive visual comparison of multiple trees. In *Proceedings of the VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011* (Oct. 2011), pp. 31–40.

[BM13] Brehmer M., Munzner T.: A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (Dec. 2013), 2376–2385.

[Bre01] Breiman L.: Random forests. *Machine Learning 45*, 1 (Oct. 2001), 5–32.

[BSKR19] Bernard J., Sessler D., Kohlhammer J., Ruddle R. A.: Using dashboard networks to visualize multiple patient histories: A design study on post-operative prostate cancer. *Transactions on Visualization and Computer Graphics 25*, 3 (2019), 1615–1628.

[CCF97] Carpendale M. S. T., Cowperthwaite D. J., Fracchia F. D.: Extending distortion viewing from 2D to 3D. *IEEE Computer Graphics and Applications 17*, 4 (1997), 42–51.

[Cha06] Charmaz K.: *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications, London, Thousand Oaks, California, 2006.

[CMMK20] Cibulski L., Mitterhofer H., May T., Kohlhammer J.: Paved: Pareto front visualization for engineering design. *Computer Graphics Forum 39* (June 2020), 405–416.

[DBD17] Dimara E., Bezerianos A., Dragicevic P.: Conceptual and methodological issues in evaluating multidimensional visualizations for decision support. *IEEE Transactions on Visualization and Computer Graphics 24* (Aug. 2017), 1.

[DK13] Dogra I., Kobti Z.: Improving prediction accuracy in agent based modeling systems under dynamic environment. In *CEC 2013: Proceedings of the 2013 IEEE Congress on Evolutionary Computation* (June 2013), pp. 2114–2121.

[DSM16] Dheenadayalan K., Srinivasaraghavan G., Muralidhara V. N.: Pruning a Random Forest by Learning a Learning Algorithm. In *Machine Learning and Data Mining in Pattern Recognition* (vol. 9729). P. Perner (Ed.). Springer International Publishing, Cham (2016), pp. 516–529.

[EBJ*21] Eirich J., Bonart J., Jäckle D., Sedlmair M., Schmid U., Fischbach K., Schreck T., Bernard J.: IRVINE: A design study on analyzing correlation patterns of electrical engines. *Transactions on Visualization and Computer Graphics 28* (July 2021).

[EJS*20] Eirich J., Jäckle D., Schreck T., Bonart J., Posegga O., Fischbach K.: VIMA: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *Proceedings of the VDS - Symposium on Visualization in Data Science at IEEE VIS* (Nov. 2020).

[EW11] van den Elzen S., van Wijk J.: BaobabView: Interactive construction and analysis of decision trees. In *Proceedings IEEE Symposium on Visual Analytics Science and Technology (VAST)* (2011), pp. 151–160.

[FI18] Fletcher S., Islam M.: Comparing sets of patterns with the jaccard index. *Australasian Journal of Information Systems 22* (Mar. 2018), 1–17.

[GRF00] Giacinto G., Roli F., Fumera G.: Design of effective multiple classifier systems by clustering of classifiers. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (2000), vol. 2, pp. 160–163.

[HH15] Hänsch R., Hellwich O.: Performance assessment and interpretation of random forests by three-dimensional visualizations. In *Proceedings of the IVAPP - Conference on Information Visualization Theory and Applications* (Jan. 2015), pp. 149–156.

[HKB*19] Heimes H., Kampker A., Bührer U., Steinberger A., Eirich J., Krotil S.: Scalable data analytics from predevelopment to large scale manufacturing. In *Proceedings of the APCORISE - 2nd Asia Pacific Conference on Research in Industrial and Systems Engineering* (Apr. 2019).

[HWWH19] Hänsch R., Wiesner P., Wendler S., Hellwich O.: Colorful trees: Visualizing random forests for analysis and interpretation. In *Proceedings of the WACV - Winter Conference on Applications of Computer Vision* (Jan. 2019), pp. 294–302.

[JFSK15] Jäckle D., Fischer F., Schreck T., Keim D.: Temporal MDS plots for analysis of multivariate data. *IEEE Transactions on Visualization and Computer Graphics 22* (Nov. 2015), 1.

[JSM*17] Jäckle D., Stoffel F., Mittelstädt S., Keim D., Reiterer H.: Interpretation of dimensionally-reduced crime data: A study with untrained domain experts. In *Proceedings of the IVAPP - Conference on Information Visualization Theory and Applications* (Jan. 2017), pp. 164–175.

[KLKR90] Kaufman L., Leonard Kaufman P., Rousseeuw P.: *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken (New Jersey): Wiley-Interscience Publication, 1990. https://books.google.de/books?id=Q5wQAQAAIAAJ.

[KvdWVW01] Kleiberg E., van de Wetering H., Van Wijk J. J.: Botanical visualization of huge hierarchies. In *INFOVIS'01: Proceedings of the INFOVIS - Symposium on Information Visualization* (USA, 2001), IEEE Computer Society, pp. 87.

[Lev66] Levenshtein V. I.: Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady 10* (Feb. 1966), 707.

[LKS*11] Landesberger T., Kuijper A., Schreck T., Kohlhammer J., Wijk J., Fekete J. D., Fellner D.: Visual analysis of large graphs: State-of-the-art and future research challenges. *Computer Graphics Forum 30* (Sep. 2011), 1719–1749.

[LS07] Liu Y., Salvendy G.: Visualization support to better comprehend and improve decision tree classification modelling process: A survey and appraisal. *Theoretical Issues in Ergonomics Science 8* (Jan. 2007), 63–92.

[LWSG13] Louppe G., Wehenkel L., Sutera A., Geurts P.: Understanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems 26* (Dec. 2013), 431–439.

[MA14] Miksch S., Aigner W.: A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics, Special Section on Visual Analytics 38* (2014), 286–290.

[Mac67] MacQueen J.: Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (Berkeley, CA, 1967), University of California Press, pp. 281–297.

[MQB19] Ming Y., Qu H., Bertini E.: Rulematrix: Visualizing and understanding classifiers with rules. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 342–352.

[MRS08] Manning C. D., Raghavan P., Schütze H.: *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[MS04] Miglio R., Soffritti G.: The comparison between classification trees through proximity measures. *Computational Statistics & Data Analysis 45* (Apr. 2004), 577–593.

[Mun10] Munzner T.: A nested model for visualization design and validation. *Visualization and Computer Graphics, IEEE Transactions on 15* (Jan. 2010), 921–928.

[NKT08] Ntoutsi I., Kalousis A., Theodoridis Y.: A general framework for estimating similarity of datasets and decision trees: Exploring semantic similarity of decision trees. In *Proceedings of the SIAM - International Conference on Data Mining* (Apr. 2008), pp. 810–821.

[PBLG19] Peres R., Barata J., Leitão P., Garcia G.: Multistage quality control using machine learning in the automotive industry. *IEEE Access 7* (June 2019), 1.

[Pea01] Pearson K.: LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2*, 11 (1901), 559–572.

[Per11] Perner P.: How to interpret decision trees? In *Proceedings of the Advances in Data Mining: Applications and Theoretical Aspects* (Aug. 2011), pp. 40–55.

[Per13] Perner P.: How to compare and interpret two learnt decision trees from the same domain? In *Proceedings of the WAINA - International Conference on Advanced Information Networking and Applications Workshops* (Mar. 2013), pp. 318–322.

[RAZ*18] Ritter C., Altenhofen C., Zeppelzauer M., Kuijper A., Schreck T., Bernard J.: Personalized visual-interactive music classification. In *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)* (2018), C. Tominski and T. von Landesberger (Eds.), The Eurographics Association.

[Sau20] Sauro J.: Measuring usability with the system usability scale (SUS). http://measuringu.com/sus [Accessed 5th November 2020].

[SBS94] Someren M., Barnard Y., Sandberg J.: *The Think Aloud Method—A Practical Guide to Modelling Cognitive Processes*. Academic Press, 1994.

[SFMB12] Sedlmair M., Frank A., Munzner T., Butz A.: RelEx: Visualization for actively changing overlay network specifications. *IEEE Transactions on Visualization and Computer Graphics 18* (Dec. 2012), 2729–2738.

[SHB*14] Sedlmair M., Heinzl C., Bruckner S., Piringer H., Moller T.: Visual parameter space analysis: A conceptual framework. *IEEE Transactions on Visualization and Computer Graphics 20* (Dec. 2014), 2161–2170.

[Shn92] Shneiderman B.: Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on Graphics 11*, 1 (Jan. 1992), 92–99.

[SIB*11] Sedlmair M., Isenberg P., Baur D., Mauerer M., Pigorsch C., Butz A.: Cardiogram: Visual analytics for automotive engineers. In *Proceedings of the Conference on Human Factors in Computing Systems* (May 2011), pp. 1727–1736.

[SJC08] Shotton J., Johnson M., Cipolla R.: Semantic texton forests for image categorization and segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition* (June 2008), pp. 1–8.

[SMF*20] Suschnigg J., Mutlu B., Fuchs A., Sabol V., Thalmann S., Schreck T.: Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Proceedings of the EDBT/ICDT - 3rd International Workshop on Big Data Visual Exploration and Analytics* (2020).

[SMM12] Sedlmair M., Meyer M., Munzner T.: Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics 18* (Dec. 2012), 2431–2440.

[SS18] Szücs D., Schmidt F.: Decision tree visualization for high-dimensional numerical data. In *Proceedings of the 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (2018), pp. 190–195.

[SSS*14] Sacha D., Stoffel A., Stoffel F., Kwon B. C., Ellis G., Keim D. A.: Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 1604–1613.

[SZS*16] Sacha D., Zhang L., Sedlmair M., Lee J., Peltonen J., Weiskopf D., North S., Keim D.: Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics 23* (Jan. 2016), 1–1.

[TD04] Tan P., Dowe D.: MML inferenc of oblique decision trees. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)* (vol. 3339) (2004), pp. 1082–1088.

[TPV09] Tsoumakas G., Partalas I., Vlahavas I.: An ensemble pruning primer. In *Applications of Supervised and Unsupervised Ensemble Methods. Studies in Computational Intelligence* (vol. 245). J. Kacprzyk, O. Okun, G. Valentini (Eds.). Springer, Berlin, Heidelberg (2009), pp. 1–13.

[Tuf90] Tufte E. R.: *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

[Via13] Viaene S.: Data scientists aren't domain experts. *IT Professional 15* (Nov. 2013), 12–17.

[WFH*01] Ware M., Frank E., Holmes G., Hall M., Witten I.: Interactive machine learning: Letting users build classifiers. *International Journal of Human-Computer Studies 55* (Sep. 2001), 281–292.

[WL17] Weinberg A., Last M.: Interpretable decision-tree induction in a big data parallel framework. *International Journal of Applied Mathematics and Computer Science 27* (Dec. 2017), 737–748.

[WL19] Weinberg A., Last M.: Selecting a representative decision tree from an ensemble of decision-tree models for fast big data classification. *Journal of Big Data 6* (Dec. 2019), 23.

[XHC*07] Xu Y., Hong W., Chen N., Li X., Liu W., Zhang T.: *Parallel Filter: A Visual Classifier Based on Parallel Coordinates and Multivariate Data Analysis* (vol. 4682). Springer, 2007, pp. 1172–1183.

[YXZC12] YANG M., XU H., ZHU D., CHEN H.: *Visualizing the Random Forest by 3D Techniques*. Springer, 2012, pp. 639–645.

[ZF07] ZHONG P., FUKUSHIMA M.: Regularized nonsmooth newton method for multi-class support vector machines. *Optimization Methods & Software 22* (Feb. 2007), 225–236.

[ZJ12] ZHANG X., JIANG S.: A splitting criteria based on similarity in decision tree learning. *Journal of Software 7* (Aug. 2012), 1775–1782.

[ZZY17] ZHANG C. X., ZHANG J. S., YIN Q. Y.: A ranking-based strategy to prune variable selection ensembles. *Knowledge.-Based Systems 125* (2017), 13–25.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information

6.3 P3: IRVINE: A Design Study on Analyzing Correlation Patterns of Electrical Engines

Full reference *Eirich, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. IEEE Transactions on Visualization and Computer Graphics, 07 2021*

IRVINE: A Design Study on Analyzing Correlation Patterns of Electrical Engines

Joscha Eirich, Jakob Bonart, Dominik Jäckle, Michael Sedlmair,
Ute Schmid, Kai Fischbach, Tobias Schreck, and Jürgen Bernard

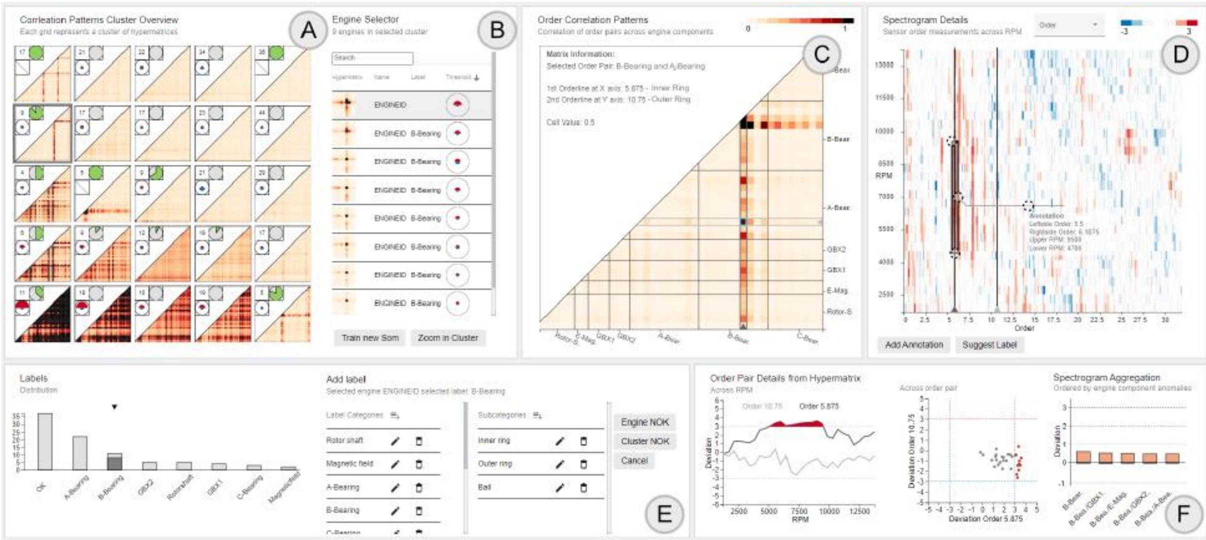


Fig. 1. The IRVINE system. Users have an overview over clusters in (A). They can select clusters in (A) and engines in (B). After selecting an engine in (B), the acoustic signature of the engine is displayed in (C) and respective raw acoustic measurements in (D). Detailed information about selections from (C) is shown as line chart and scatter-plot and bar chart in (F). After the analysis of an engine, the user can assign a label in (E) and provide an annotation for the label in (D).

Abstract—In this design study, we present IRVINE, a Visual Analytics (VA) system, which facilitates the analysis of acoustic data to detect and understand previously unknown errors in the manufacturing of electrical engines. In serial manufacturing processes, signatures from acoustic data provide valuable information on how the relationship between multiple produced engines serves to detect and understand previously unknown errors. To analyze such signatures, IRVINE leverages interactive clustering and data labeling techniques, allowing users to analyze clusters of engines with similar signatures, drill down to groups of engines, and select an engine of interest. Furthermore, IRVINE allows to assign labels to engines and clusters and annotate the cause of an error in the acoustic raw measurement of an engine. Since labels and annotations represent valuable knowledge, they are conserved in a knowledge database to be available for other stakeholders. We contribute a design study, where we developed IRVINE in four main iterations with engineers from a company in the automotive sector. To validate IRVINE, we conducted a field study with six domain experts. Our results suggest a high usability and usefulness of IRVINE as part of the improvement of a real-world manufacturing process. Specifically, with IRVINE domain experts were able to label and annotate produced electrical engines more than 30% faster.

Keywords: Design study, interactive labeling, interactive clustering.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces (GUI); User-centered design

1 Introduction

The automotive industry is currently in the midst of its greatest change in the last 100 years, namely in the direction of electromobility [28]. With the setup of electrical engines (from here on: engines), of course, new, previously unknown errors are introduced. For car manufacturers, it is extremely important to identify and understand these errors to meet high-quality standards. However, detecting and understanding such errors is currently hampered by two major challenges: **(1)** Serial manufacturing processes produce a large quantity of parts, hence, resulting in a vast amount of recorded sensor data during testing. On-site interviews with test engineers showed that they are required to analyze more than forty interdependent signals from hundreds of produced engines manually and per hour. Furthermore, the measured signals can be related to multiple sub-components of an engine. This situation poses a major challenge and as a result test engineers are capable to

- *Joscha Eirich, Ute Schmid, and Kai Fischbach are with University of Bamberg. Email: firstname.lastname@uni-bamberg.de*
- *Jakob Bonart is with Fraunhofer IWU. Emai: jakob.bonart@bmw.de*
- *Dominik Jäckle, Joscha Eirich, and Jakob Bonart are with BMW Group. Email: firstname.lastname@bmw.de*
- *Michael Sedlmair is with University of Stuttgart. Email: Michael.Sedlmair@visus.uni-stuttgart.de*
- *Tobias Schreck is with Graz University of Technology. Email: tobias.schreck@cgv.tugraz.at*
- *Jürgen Bernard is with University of Zurich. Email: bernard@ifi.uzh.ch*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

only analyze few parts in detail in the given time. **(2)** Because only a few engines can be analyzed and their errors classified in detail, the gained knowledge is particularly precious for the overall improvement of the testing procedures. Yet, it is currently unknown how the gained knowledge can be stored or transferred. To address the named challenges, we present *IRVINE (InteRactiVe cluseterINg labEling)*, a Visual Analytics (VA) system that results from a design study [49] project carried out together with automotive engineers at BMW.

We base our study on acoustic data that is collected through the propagation of sound inside the engines. The analysis of acoustic data allows specific errors to be recognized. For example, a dirty bearing sounds differently from a clean bearing. The interaction between different acoustic frequencies forms so-called signatures. Signatures typically comprise one primary and multiple secondary error symptoms. On top of raw acoustic data, we also include the derived correlations in acoustic signatures, which enable interdependent error analyses. Given the collected acoustic data, *IRVINE* provides capabilities for clustering similar signature correlations. To gain a fine-grained overview of similar signatures, clustering can be performed interactively by engineers on a subset of clusters or engines with previously provided labels. Engineers can then select specific engines of a cluster for an in-depth analysis. To record their findings we introduce knowledge capturing capabilities through labeling and annotation. In this regard, we refer to labeling as the error class (e.g. B-Bearing error), which serves as a basis for the classification of further errors. In contrast, we refer to annotations as a means to record the reason for an error within the raw sensor data (e.g. Threshold violation in B-Bearing measurement).

To better support the goals and tasks of engineers, we first familiarized with their domain and gained an understanding of their problems. The *IRVINE* system was iteratively designed, developed, and evaluated in close collaboration with engineers at BMW. The overall goal of the system is to support engineers in the detection and analysis of error-prone produced engines. *IRVINE* was hereby tailored to fit their domain-specific requirements and designed to be seamlessly integrated into their daily work. As a result, the systems can handle vast amounts of data enabling engineers to easily analyze multiple engines, their signatures, and acoustic raw data. Labels and annotations are stored in a knowledge base, which is used to both enhance *IRVINE* and to be available for different stakeholders, who can benefit from labels and annotations.

In summary, our contributions are: (1) The problem characterization and abstraction of the studied use case; (2) the reporting of the interactive design of IRVINE; and (3) the evaluation of IRVINE together with six automotive engineers and reflections of our design process.

2 Related Work

We start by giving a brief summary on related work about *interactive clustering* (Section 2.1), *interactive labeling* (Section 2.2), and previous *design studies in the automotive sector* (Section 2.3).

2.1 Visual Interactive Clustering

Interactive clustering approaches have been proposed for various data types, including trajectory data [3], text documents [40], (social) networks [37], and time series data [8]. Furthermore, other visualization approaches focus specifically on the visual analysis of sensor data [1] including a variety of analysis goals, such as segmentation, clustering, or classification [38]. In scenarios where very large data sets are analyzed, some approaches make use of matrix-based visualizations, for instance to group similar matrices that show changes in brain connectivity networks [4]. With respect to our clustering scenario, we focus on the grouping of large amounts of acoustic signatures of engines on their similarities, with the help of matrix visualizations. Some approaches explicitly leverage domain knowledge, e.g. Yang et al. [59], focusing on interactive steering methods to visually constrained clustering using both user and publicly available knowledge. The selection of specific subsets of interest where another subsequent computation of the clustering is applied plays also an important role [42].

Interactive clustering is often combined with dimensionality reduction techniques [43] to benefit from the complementary strengths of a) reducing the number of instances from many to a few (via clustering) and b) reducing multiple data dimensions to a low-dimensional visual representation (dimensionality reduction). We use the Self-Organizing Maps (SOM) [32] clustering algorithm, which naturally combines both steps. Vesanto was one of the pioneers who showed the benefits of SOMs for visual cluster analysis [55], and many methodological contributions to interactive clustering followed from VA research. Schreck et al. [46] presented a VA system that allowed the interactive visual initialization, quality assessment, and refinement of SOMs. *IRVINE* also supports the interactive refinement of SOMs [41], with both with different parameters and data subsets of interest. The SOM has also been used for the visual analysis of sensor data, in MotionExplorer [12] to facilitate exploratory search and in FuryExplorer [57] for the comparison of cluster patterns with attached metadata. Finally, the *SOMFlow* system serves as a platform for the interactive exploration of multiple SOMs, the analysis and refinement of intermediate results, and the back-and-forth navigation along the analytical workflow [42].

In our design study, we use interactive clustering to support the engineers in the analysis of acoustic data. Particularly inspiring for the specific design of *IRVINE* are the iterative training and refinement, the training of subsets of interest, and the analysis of individual clusters and cluster elements in detail for downstream labeling tasks.

2.2 Visual Interactive Labeling

As well as interactive clustering, labeling is also a frequently supported task in VA [20]. In this regard, Bernard et al. [13] propose the concept of *Visual-Interactive-Labeling* (VIAL), to label yet unknown data in an interactive exploratory setup. VIAL hereby bridges the gap between active learning [50] and advanced visualization concepts, where the type of labels depends on the given task and approach. Categorical labels count to the more common types, which can either be of binary or multi-valued nature [9]. While binary labels would allow simple user feedback, such as "ok vs. not ok", *IRVINE* supports multi-valued labels, enabling a more specific tagging of different classes for a data instance. Other examples for systems that enable categorical labeling are provided for textual documents [25], bio images [7], handwritten digits [9], or video streams [27]. Considering our use case, we focus on the assignment of categorical multi-valued labels.

A second important type of labels is continuous labels, which are often applied when a more fine-grained degree of interestingness is required. Here, systems exist for candidate rating and evaluation [56] or to choose between irrelevant and relevant views [6]. Yet other labeling approaches allow the comparison of pairs of objects [10] or groups of objects [16] in combination with algorithmic models using this implicit feedback to adjust the attribute weightings or the feature space. Finally, another type of user feedback is to annotate features or data attributes directly. For instance, approaches exist that support the dynamic evaluation of feature subsets and resulting models [60] or to annotate images while relating model results to their input features [34]. With respect to our use case, engineers will be able to annotate local regions of interest within the analyzed feature space.

Many of the labeling and annotation approaches were built to train some kind of more or less explainable machine learning model, such as decision trees [44], or support vector machines [25] to name a few. However, especially when users are not familiar with machine learning, previously labeled and annotated data instances, may already provide valuable information to guide analyses. Thus instead of creating a "black-box" classifier [39], we will focus on the storage and availability of labels and annotations to guide user analyses.

2.3 Design Studies in the Manufacturing Sector

Design studies in the automotive sector are mostly carried out for engineering design and anomaly detection. In this context, efforts were carried out to visualize the exploration of multi-criteria alternatives for rotor designs [19], in-car communication networks [47, 48], or anomaly detection with test stations from large scale manufacturing processes [21, 53]. While some of the mentioned studies acknowledge the need for storing expert knowledge [21], the only systems we found which addresses the problem of specifically leveraging that kind of knowledge is *Cardiogram* [48]. *Cardiogram* addresses the problem of debugging masses of traces from in-car communications networks to become error-free. In turn, the problem at hand which *IRVINE* addresses is the analysis of acoustic data, which necessitates different kinds of visualization approaches. Thus, grounded on previous findings, we did build a system that allows domain experts to externalize their knowledge from the analysis of acoustic data, supporting them in their high cognition task of analyzing engines.

3 METHODS

During this study, we primarily followed Sedlmair et al.'s nine-stage framework for design studies [49]. In addition, we used the *Nested Model* for visualization design and validation by Munzner [35]. The nested model guides a more detailed problem characterization, the data operation and abstraction, the visual encoding and interaction design, and the algorithm design.

Our system development went through four main iterations, during which we interviewed engineers, tested design alternatives, and held critical discussions with visualization experts. Each iteration was carried out in close collaboration with one engineer at BMW with extensive experience in the optimization and design of test procedures for *engines*. He accompanied the system development with knowledge about the problem domain (Section 4.2), the data abstraction (Section 4.3), and resulting tasks (Section 4.4). He also gave a constant stream of feedback on the visual design of our system. The exchange with the engineer took place on up to four meetings per week ranging from 30-60 minutes. During these meetings, fundamental characterization and design aspects were discussed and open issues clarified, while he provided us with feedback about relevant tasks and the system design. An evaluation of the system was carried out with four engineers at BMW. Based on the feedback from the evaluation, we refined *IRVINE*, which resulted in increased labeling and annotation speed with two more engineers. Methodological details for this downstream evaluation are provided in Section 8.

4 ABSTRACTIONS

We report the characterization of the problem domain in the field of quality control in the manufacturing sector. First, we provide an introduction to automated part testing in automotive engineering (Section 4.1), followed by a description of the domain of our collaborators (Section 4.2). Finally, we report on the task abstraction, supporting engineers with the analysis of error-prone engines (Section 4.4).

4.1 The Automotive Engineering Domain

The manufacturing of engines requires testing each engine in various stations along the assembly line. To that end, engineers analyze acoustic data by recording the noises of engines when being simulated with real-world conditions on a test bench. With *IRVINE*, we build upon the common case of acceleration tests, where a speed ramp is running from 2500 to 14000 revolutions per minute (rpm) to capture a sonic image of the engine at all relevant rotations. The measurement of these acoustic measurements enables the detailed analysis of engine sub-components (e.g. the gear or the bearing) and allows to allocate the primary symptom of an error (e.g. a gear with a scratched surface sounds different from an undamaged gear). According to the measurements of acoustic data, engineers can exactly identify the sub-components, which are known a priori from lab experiments carried out at earlier development stages.

What is challenging for engineers though, is the existence of secondary error mechanisms accompanying each specific primary error. In fact, fault mechanisms are visible in the primary error source, but also partially in their secondary error sources. The engineers reported that a good understanding of the interplay between primary and secondary errors sources is extremely valuable to further improve automated part testing with acoustic measurements. At BMW, engineers have developed a procedure for the analysis of relations between primary and secondary error symptoms, based on the systematic *correlations* of vast amounts of pairs of measurement values. Aggregated forms of these correlation results are called the **signature** of an engine. A detailed description of the signature computation process is outlined in Section 5.1.

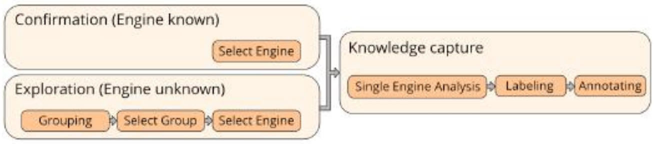


Fig. 2. Data analysis workflow of engineers. Engineers are either confirmers or explorers with the shared goal to select, analyze, label, and annotate a single engine.

Based on the observation of engineers and interviews conducted in the early stages of the project, we abstracted the principal workflow of engineers for the analysis of engines through acoustic data. In general, engineers switch between two main goals referring to (1) identifying engines of interest as well as (2) analyzing, labeling, and annotating engines. The workflow is shown in Figure 2, offering interesting nuances with respect to engine identification. Currently, the dominating information-seeking behavior [29, 51] of engineers is of confirmatory nature. Confirmers already know engines of interest, aiming for the validation of hypotheses about errors in an engine. In addition, the characterization of the domain problem in Section 4.2 shows the need for exploratory analysis support. Explorers are open to large varieties of engines and respective signatures, aiming at formulating new hypotheses about unknown engines and signatures. The workflow for unknown engines roughly ranges from multiple engines, over a small selection of engines, down to single engines of interest. Both explorers and confirmers share the same goal downstream: engineers review the signature of an engine in detail to allocate symptoms and compare it to the raw acoustic measurements to validate these symptoms. As a result, a label can be assigned to the engine. Finally, engineers also make annotations directly in the acoustic raw data to externalize knowledge about the source of an error.

4.2 Domain Problem

Both the confirmation and the exploration of engines as outlined in Figure 2 are time-consuming and tedious processes due to the following two main problems: First, large amounts of available signatures and recorded acoustic raw data measurements make it almost impossible for engineers to keep track of all engines at a desired level of detail. As a result, the detailed analysis, labeling, and annotation of engines can only be performed in an anecdotal manner. Only the confirmation of specific engines is supported, while an exploration of unknown signatures from unknown engines remains an open issue. Here, engineers reported that above all the efficient grouping of similar signatures from engines is not possible at the moment. The systematic analysis of all engines and error types falls short. Second, the knowledge, which is created during the detailed analysis of engines has a high impact on the improvement of automated part testing procedures. For engineers, it is especially useful to know, which error was observed in engines and where the source of the error can be seen in the data. However, occurred errors in engines are labeled too seldom by engineers, while the annotation of the cause of an error is not supported at all.

4.3 Data Abstraction

Engineers record their data inside a test bench with a sensor for measuring the noise of the engine. For that purpose, the engine's rpm is steered under controlled conditions to analyze the behavior of engines. Measuring the noise allows the evaluation of the acoustic properties of components and their technical condition. The result of such acoustic measurements is a three-dimensional data structure consisting of *loudness* measured across all possible combinations of *rpm* and *orders*. To shift focus on anomalies, the loudness values are often replaced by *residuals*, e.g., by the deviation from measured values to mean values μ of an ensemble of engines (expected value), divided by the standard deviation σ (see also eq. (1)) A 2D pixel-based visualization of these measurements often used by engineers is the **spectrogram**, as shown in Figure 3.

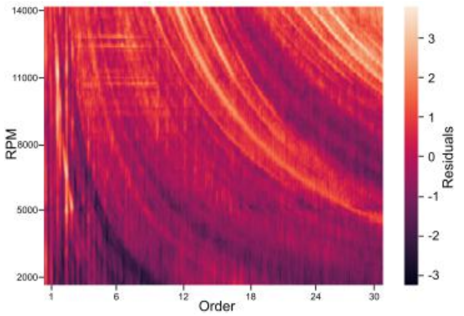


Fig. 3. Spectrogram of acoustic measurements for an engine. Rpm is shown on the y- and orders on the x-axis. The color indicates the volume of the measurement compared to the mean distribution of engines.

The **rpm** value is displayed on the y-axis, describing the acceleration of the engine up to a maximum speed. The **order** is displayed on the x-axis, which is the relation between a measured frequency and the speed of the engine during measurement, consequently describing how often an excitation occurs per revolution. Orders are beneficial, as they allow the fine-grained analysis of sub-components through an expert's eye. By analyzing the geometry of the rotating engine, engineers are able to derive the measured sub-component of an engine up to a very detailed level (e.g. 24th teeth of a gear). **Residuals** are shown on a color scale, where brighter colors represent particularly loud orders. According to the engineers, in serial manufacturing regions of loud frequencies are a good indicator for different types of errors. The three-dimensional measurement data forms the basic representation of an engine through abstract data, providing the raw data for any analysis scenario on engine errors in IRVINE.

Figure 3 shows a residual order-spectrogram. Specifically, 512 order lines (each column in Figure 3) ranging from 2500 - 14000 rpm are recorded for each engine. To identify signatures in order combinations that might result in a faulty engine, theoretically every possible order combination has to be analyzed manually for each engine. For engineers, this would result in 262,144 possible combinations, which is not feasible for a manual analysis. Hence, the engineers narrowed down the orders to the 41 most relevant ones. Based on the informed selection of orders, they can be connected to the seven main sub-components of an engine (*rotor-shaft-, electromagnetic-, first gear-, second gear-, A-bearing-, B-bearing-, and C-bearing orders*). Knowing the connected sub-components for individual orders allows the fine-grained analysis and labeling of engine errors at a sub-component level.

4.4 Task Abstraction

The intensive collaboration with engineers at BMW helped us to understand the domain, the domain problem, and the desired workflow for the analysis.In the following, we present the task abstraction that will support engineers in reaching their goals. Likewise, these tasks will serve as primary design targets for IRVINE.

$\mathbf{T}_1$ *Gain overview of engines*: By taking the role of explorers, engineers need a structured overview of the data, in our case of engines represented by their acoustic signals. Grouping engines by the similarity of acoustic signals using clustering algorithms would be desirable. In addition, interactive grouping to adapt to the information need of individual engineers would be useful, e.g., based on different steering parameters (such as the number of groups) or filtered subsets of engines.

$\mathbf{T}_2$ *Drill-down to engines*: Engineers exploring large numbers of engines need support for the drill-down to engines of interest. The information need may differ between clusters of engines, the labeling status, or the degree of the anomaly of engines.

$\mathbf{T}_3$ *Identify engine of interest*: The workflow of both explorers and confirmers includes the identification of single engines as an entry point into an in-depth analysis. Identification may be through knowledge about a particular engine, by special or even unique acoustic signatures, or by traversing several (similar) engines in the exploration process.

$\mathbf{T}_4$ *Analyze single engine*: Engineers need to analyze individual engines in detail for being able to assign labels on a profound basis. Single engines are both represented by their acoustic signature and the raw acoustic measurements. The analysis is also supported by stored domain knowledge from the systems knowledge base.

$\mathbf{T}_5$ *Assign label to engine*: Engineers need to assign labels of error categories. In close collaboration, we formed a default alphabet of label categories as follows: *Error/Electromagnetic-Field*, *Error/First Gear*, *Error/Second Gear*, *Error/A-Bearing*, *Error/B-Bearing*, *Error/C-Bearing*, *No error*. In addition, engineers also expressed the need to leave the label alphabet open for modifications, as knowledge about error variations will constantly grow as IRVINE is used. Finally, to enhance efficiency in combination with $\mathbf{T}_2$, it would be desirable to label single engines but also multiple similar engines at a glance.

$\mathbf{T}_6$ *Annotate acoustic measurements*: The cause of a labeled engine can be annotated by marking the respective region inside of the acoustic raw data of an engine (the spectrogram) serving two purposes. First, annotations can be used to review how similar labels were annotated by different users. Second, a sufficient amount of annotations for a given label will allow building thresholds for semi-automatic error detection.

5 Non-visual data analysis support

In this section, we present the measurement of acoustic data and pre-processing steps, carried out to prepare the data. Based on the data abstraction (Section 4.3), we describe the computation of an engines signature (Section 5.1). Next, we describe details on the feature extraction and model training for our clustering approach (Section 5.2).

5.1 The Computation of Signatures

To analyze relations between primary and secondary symptoms of errors, engineers calculate correlations in-between spectrograms. The motivation is given in Section 4.3 and the schematic process for each engine is outlined in Figure 4. An exemplary process is described for a single signal pair (signal A and B) and one engine to increase clarity. This process is consequently applied to all available signals and engines.

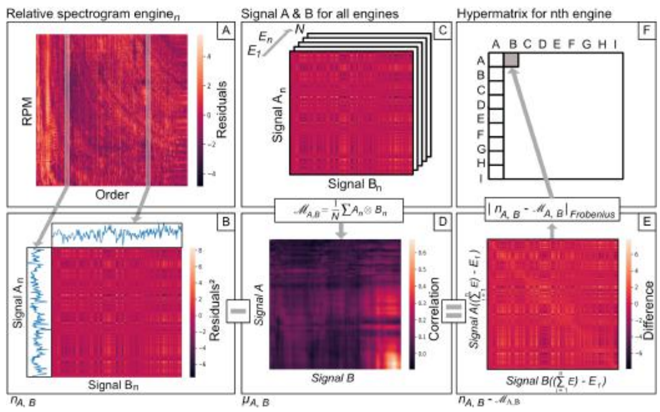


Fig. 4. Computation of our Hypermatrix. From a given spectrogram in (A), we extract two columns that are correlated with each other in (B). This is done for each signal pair over all engines in (C). Next, we subtract the resulting mean signal combination in (D) from each signal combination of a single engine which results in the deviation of a signal pair from one engine to all other engines in (E). Each signal pair is then aggregated and stored in a new matrix in (F).

As a first step, we calculate the mean and standard deviation for the ensemble of engines at hand. Consequently, the relative deviation of the i-th engine to the mean μ for each (rpm, order)-tuple is then expressed in units of the standard deviation σ:

$$\mathrm{Residual}(ord, rpm)_i = \frac{\mathrm{value}(ord, rpm)_i - \mu(ord, rpm)}{\sigma(ord, rpm)}. \quad (1)$$

An example of the resulting residual spectrogram described in Section 4.3 is given in (A). Next, we extract two measurements A and B (being 1D- curves each), and calculate the outer product for the pair, resulting in a 2D-matrix (see (B)). The choice of possible pairs is restricted to 41 relevant orders from the data abstraction, which are used to derive relevant order combinations. Calculating the mean value of the 2D-matrices over all engines for each entry (C) results in the correlation matrix (D). This correlation matrix effectively consists of Pearson correlations for pairs of rpm-values of two extracted orders. Therefore, the resolution regarding different engine speeds and the corresponding orders is still retained. To extract the difference in the correlation, we subtract the correlation matrix from each outer product resulting in a matrix describing correlations inside each measurement (E). This matrix is then reduced onto its cumulated deviation using the Frobenius-norm. Consequently, each cumulated deviation for each pair of orders is then ordered into the reduced-Difference-Correlation-Matrix, which we call **Hypermatrix**.

Figure 5 shows three exemplary *Hypermatrices* of different engines and their corresponding signature patterns. The first belongs to an *OK* engine with no errors. The second represents an engine with an anomaly in the B-Bearing but no error and the third an error in the B-Bearing. Each matrix was build with 41 signals, provided by our domain expert, which can be related to one of an engine's sub-components as described in the data abstraction. The region for each sub-component was marked by our domain expert with black lines. Hence, secondary error sources in sub-components can be identified by reviewing correlations in each region. Especially, in the second and third *Hypermatrix* one can clearly see which order is the primary error source in the signature and which orders resonate with the excitation thus representing secondary error sources.

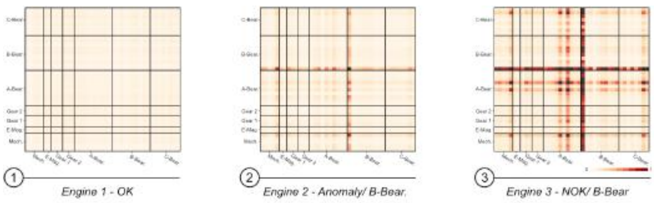


Fig. 5. Hypermatrices showing acoustic signatures. (1) with no error, (2) an anomaly in the B-Bearing, and (3) an error in the B-Bering.

5.2 Feature Engineering and Model Training

To group similar acoustic signatures, we extract features from each engine's *Hypermatrix*. The feature extraction process is shown in Figure 6. The combination of all seven sub-components results in 28 possible combinations and is depicted as Region (*R*) in Figure 6. From each *R*, the sub-matrix is extracted. After experimenting with different feature sets, we received the best clustering results extracting the *maximum* of each of the 28 resulting matrices. This results in a 28x1 feature vector.

The engineers agreed on the usefulness of this approach, as in their application domain, louder noises tend to be the cause for an error. However, different features can also be used as input by applying minor changes to the feature extraction process.

To cluster similar *Hypermatrices*, the 28x1 feature vector is used as input for a Self-Organizing Map (SOM) [32], as it nicely combines clustering with dimensionality reduction functionality. For the computation of the SOM, we follow a simplified version of the standard training process described by Kohonen [31]. We set the SOM grid size such as to expect at least one data vector per node, which accounts for very specific error types. We initialize the SOM prototype vector dimensions with random numbers between 0 and 1 and train the SOM by iterating over the input data vectors and adjusting the SOM nodes. Specifically, we find for each input data vector the best matching SOM prototype unit (BMU) according to *Euclidean* distance. We then adjust the BMU and its neighborhood according to a linearly decreasing learning rate and circular neighborhood kernel. This configuration comprises the initial implementation of our SOM and can be changed by the user as described in detail in Section 6.2. We note that our heuristic setting of parameters already gave us robust results for our application, hence, we did not see the need for parameter optimizations. In principle, also other visual clustering techniques may be applied besides SOM. We particularly chose SOM because of its robustness in our application domain, and as it gives an overlap-free rectangular layout that well supports visual comparison tasks.

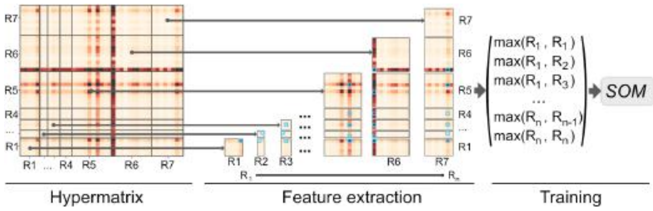


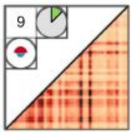
Fig. 6. Feature engineering process. All sub-component combinations in the in the *Hypermatrix* form 28 Regions (R) of interest. From each R, the sub-matrix is extracted and from each sub-matrix, the maximum is computed. All maximas of regions form the input vector for the SOM.

6 The IRVINE System

Figure 1 shows an overview of the main views of *IRVINE*, as also outlined in Section 6.1. Each view is marked from A-E, where we will use this notation hereafter to refer to *IRVINE's* individual views. Due to non-disclosure agreements with BMW, we are not able to show engine ids.

6.1 Overview of the System

In (A), all clusters of similar *Hypermatrices* as outlined in the Sections 5.1 and 5.2 are displayed. Each *Hypermtarix* in the grid is represented by the mean aggregation of all *Hypermatrices* in a cluster of engines. The grid view serves to get an overview over all clusters and their individual properties ($\mathbf{T}_1$) and to support the decision of which cluster to select ($\mathbf{T}_2$). To visually encode a *Hypermatrix*, we use a sequential color scale. This is appropriate since all *Hypermatrix* values range from 0-1, where 1 represents a high correlation between a pair of sub-components and 0 a low correlation.



Each grid in the cluster view also contains three additional rectangles. The first shows the number of engines in a cluster. The second represents the number of already labeled engines as a pie chart, where green are labeled engines and grey not labeled ones. The aggregated deviation of all engines in a cluster to the serial distribution of signatures from all available engines is displayed as two colored glyphs. Here, red arcs represent engines, which contain a deviation greater than zero and blue lower zero. This separation is important because engines that are deviating upwards are louder and downwards are quieter. Hence, users get an immediate overview of how many engines are in each cluster, how many engines are already labeled, and which cluster contains the most anomalous engines. In Figure 1 the selected cluster is shown by its grey stroke.

(B) shows engines in a cluster, the *Hypermatrix* of each engine as small multiple, and the aggregated deviation of the signature of a single engine to all other engines. This view is designed to get an overview of the engine's properties and compare them to other engines in the cluster. Thus, it supports the user in selecting an engine ($\mathbf{T}_3$). If the user has a specific engine id of interest the engine list can also be filtered accordingly and thus allows for confirmation as shown in Figure 2. The *Hypermatrix* of each engine is represented in the engine list view as a small multiple, while the same glyph representation as in the cluster view is used. Initially, the list is sorted according to the euclidean distance of engines to their centroids in each cluster. However, the user can sort the list according to the deviation of engines greater than zero. In Figure 1 the selected cluster contains 9 engines, sorted according to their anomaly score and contains similar *Hypermatrices*. By clicking on "Zoom in Cluster" a detailed overview of all *Hypermatrices* in the cluster replaces the view in (A), which is demonstrated in Figure 8-1.

Of course, users have the option to return to the initial view. The user can also retrain clusters by clicking on “Train new SOM”. The process to train clusters is outlined in detail in Section 6.2.

In (C), the detailed *Hypermatrix* of a selected engine is shown. Additional information about single selected cells in the matrix is displayed in the upper left triangle. This view is designed to support the user in the analysis of a selected engine ($\mathbf{T}_4$). A matrix representation is adequate to represent the relation between pairs of sub-components (e.g. Gear and rotor shaft). As pointed out in Section 5.1, the regions of sub-components in the *Hypermatrix* are marked with additional black lines. The same color scheme as in the cluster view for *Hypermatrices* is applied. The selection of a cell in the *Hypermatrix* is supported by additional lines and triangles (dark and light grey) for each axis.

(D) visualizes a spectrogram of the selected engine according to Section 4.3. This type of visualization is the same one, engineers analyze during their daily routines and is used to support the analysis of a single engine ($\mathbf{T}_4$). Here, a diverging color scale is used. Blueish colors represent acoustic measurements, which are more quiet compared to all other engines, and reddish colors louder ones. This kind of color scale is appropriate since all values spread around 0 and deviate in different directions. By hovering over a cell in (C) the according orders of a pair of sub-components are displayed in the spectrogram. In the example in Figure 1, the sub-component pair (B-Bearing and A-Bearing) is selected. The former is shown with a triangle in dark grey and the latter with light grey. An engine can be annotated by clicking on “Add Annotation” ($\mathbf{T}_6$), which is outlined in detail in Section 6.4

(E) shows the distribution of already assigned labels as a bar chart. The main purpose of the view is to support the labeling of (selections of) engines ($\mathbf{T}_5$), as outlined in detail in Section 6.3. To show label distributions, a bar chart is an obvious choice. It would have been possible to use a pie chart, but for the labels that would have broken the guideline that there should be no more than six segments [24]. In the example in Figure 1, all but one engine in the cluster are labeled as B-Bearing error.

(F) shows three different views. First, the line chart displays two orders from the spectrogram across their rpm values. Second, the scatterplot shows the correlation of the selected order pair. Third, the bar chart shows the aggregated deviations for a region in the *Hypermatrix* to all other regions as indicated in Figure 6. These views are designed to facilitate the analysis of an engine ($\mathbf{T}_4$), where their input data are displayed when hovering over a cell in (C). In our application domain, deviations above and below three tend to be reasons for errors in the selected part. Thus, lines above and below this limit are marked as red for (+3) and blue for (-3). The purpose of the scatterplot is to provide additional information on how the two selected order lines correlate with each other. An additional overview about the five most deviating sub-component pairs in (B) is provided as a bar chart view in (E). To be consistent with our use of colors, red bars represent aggregated deviations greater than zero and blue ones lower zero.

6.2 Interactive Clustering

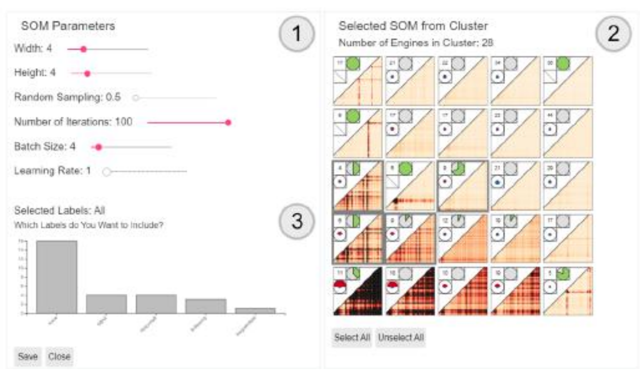


Fig. 7. Dialog to train SOM. Users can select SOM parameters in (1), choose input clusters in (2), and filter for relevant labels in (3).

To facilitate and speed up labeling, users can perform clustering based on a subset of engines in (A) - “Train new SOM”. Figure 7 shows the dialog window with the three possible options for user interaction marked from 1-3. In (1), the user can define the input parameters to train a new SOM, such as batch size or learning rate.

As input data, the user can filter all engines based on their cluster (2) or label (3). In (2), the user selects engines by clicking on a cluster in the grid to form a subset of new engines. In (3), the user selects relevant labels of interest by clicking on a bar in the bar chart. After the parameterization, the SOM is trained as described in Section 5.2 and replaces (A) in Figure 1. However, users always can return the initial visualization.

6.3 Interactive Labeling

After an analysis is complete, the user can assign a label to the component in (E) with the two list views ($\mathbf{T}_5$). Labels can also have subcategories (e.g. *“B-Bearing/Inner ring”*). The user is able to create new categories and subcategories or update and delete them in (E). After a label is selected, users can either label a single engine or the entire cluster. However, in our design study, we experienced cluster labeling only for very small *not-OK* or large *OK* clusters. Provided labels serve for the retraining of the clustering as shown in Section 6.2 and are stored in the system’s database to be available to other engineers. Entered labels further support three tasks. First, they give an additional overview over groups of engines ($\mathbf{T}_1$), because clusters which are already completely labeled are less interesting for an analysis. Second, they support the selection of an engine ($\mathbf{T}_3$), since engines that contain a label are also less probable to be selected. Third, they help in the analysis of a selected engine ($\mathbf{T}_4$). This is because they are immediately displayed in the engine list view in (B) and thus give hints on the probability of a label for the selected engine. If for example, 4 out of 5 engines in a cluster contain the same label, it is probable that the last engine also contains the same error and thus should be labeled equally.

6.4 Annotation of Sensor Data

When an engine is labeled, the cause of an error can further be annotated by the user in the spectrogram in (D). Here, the user can select the specific region in the spectrogram by a rectangle selection as shown in Figure 8-2 and -3. By dragging the edges of the rectangle, users are provided with feedback in a tooltip to which order the left and right side and to which rpm value the upper and lower side of the rectangle belongs. As well as labels, annotations can support analyses in multiple ways.

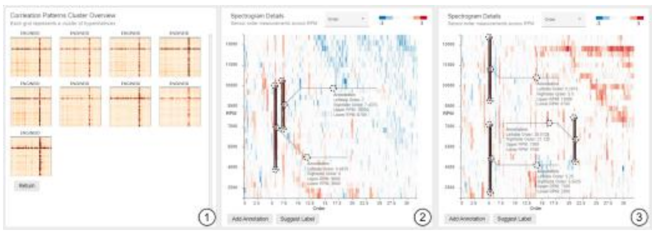


Fig. 8. (1) Drill-down of Hypermatrices from the selected cluster in Figure 1(A) and errors in the A-Bearing in (2) and a B-Bearing in (3) with according annotations.

First, users are able to review annotations from previous analyses of other engineers. By selecting a label from the bar chart view in Figure 1-E all other annotations for the same label are displayed in the spectrogram. This supports the user in the analysis of an engine ($\mathbf{T}_4$).

Second, when an engine was annotated, the user can request a label and annotation suggestion from the system. Here, all labels and respective annotations are queried from the database. Since each annotation exactly specifies the range of columns (Left Order/ right Order) and rows (Upper rpm/ Lower rpm), it allows to narrow down the search space for an anomaly in the spectrogram. Next, the three annotations and according labels with the highest threshold violations are displayed to the user. If there are no violations, the system suggests that the selected engine seems to be *OK*. We choose to give the user only the opportunity to request an annotation from the system after an annotation input was manually made by the user. With this design choice we intent to avoid blind trust in a system recommendation and thus a decision bias. The more annotations are made, the better the suggestion of further annotations becomes. All labels and annotations are stored in the system's database and can be used for different purposes, such as model training. Thus, a continuous stream of knowledge from domain experts is stored [36] with *IRVINE* via labeling and annotating.

6.5 Implementation

The system is a single-page web application written in Typescript, HTML, and CSS using the framework *Angular Js*. All views are based on *D3.js* [14]. To improve rendering speed, all SVGs are rendered as canvas elements. *IRVINE* runs on a Docker Container on a virtual machine, so that each employee inside the BMW network can access it via a public URL. The input data is processed in a separate Python application and stored in a SQL Database. All API calls run on a separate Python Flask application hosted on a virtual machine from BMW.

7 Usage Scenarios

This section outlines two usage scenarios for the interactive clustering and labeling with *IRVINE*, in line with the groups of required functionalities of Section 4.4. The first scenario (Section 7.1) shows how a new clustering can be performed to retrieve a group of similar signatures. The goal hereby lies in the fast detection of interesting signatures for labeling. The second scenario (Section 7.2) shows how users perform a detailed analysis of an engine, provide a label for a B-Bearing error, and annotate the respective acoustic measurement.

7.1 Scenario 1: Interactive Clustering

The engineer Alexandra is an explorer. For her analysis, she has 434 engines available, all of which are unknown to her at the start. Her goal is to assign labels to unlabeled engines ($\mathbf{T}_5$). She starts IRVINE, leading to the analysis state as shown in Figure 1 with a SOM clustering with 5x5 grid cells. Alexandra is interested in engines with a clear signature. She selects four grid cells with an overall of 28 engines and four different labels, namely the second gear (GBX2), the rotor shaft, the B-Bearing, and the magnetic field, as shown in Figure 7-3. To gain an overview of the characteristics of only these 28 engines ($\mathbf{T}_1$), she decides to re-train clustering with 4x4 cells only using these engines. Due to the small number of selected engines, she chooses a relatively small batch size, while for the other parameters she keeps the default settings.

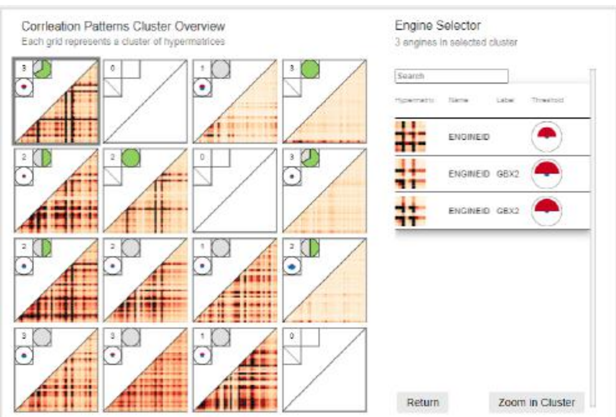


Fig. 9. Clusters resulting from a SOM training with 28 engines. The selected cluster contains three engines, where two are already labeled with an error in the second gear.

Alexandra analyzes the clustering result as shown in Figure 9 and realizes that, as she expected, all clusters contain a small number of engines between zero and three engines. In this way, she can focus on very small subsets of engines, when browsing through individual grid cells. At a glance, she identifies many different signatures, while stronger errors (with a high degree of red colors) align on the left of the SOM. Interestingly, the two clusters with the most apparent signatures were already labeled by other engineers before.

She decides to browse further and selects the top left cluster ($\mathbf{T}_2$), containing engines with well-distinguishable signatures. Two of the three engines already contain labels, hinting at errors in the second gear. The *Hypermatrix* of the unlabeled engine in the engine list view in Figure 9 is very similar to the labeled engines. Alexandra thus comes to the conclusion to label this engine also with an error in the second gear ($\mathbf{T}_5$). The proper validation of an assigned label requires a detailed analysis of the selected engine. This scenario is outlined in Section 7.2.

7.2 Scenario 2: Interactive Labeling and Annotating

The engineer Thomas is a confirmer. He uses the same data set as Alexandra but wants to analyze engines with a specific signature he discovered in previous analyses ($\mathbf{T}_4$). The cluster with the interesting signature is outlined in Figure 1 with a grey stroke in (A). For this cluster, only one engine has not been labeled yet, which is why Thomas selects this engine for a detailed analysis ($\mathbf{T}_3$). Being a confirmer, Thomas has a hypothesis that *an error in the B-Bearing can also be seen in the acoustic measurement of an A-Bearing*. By reviewing the *Hypermatrix* of the selected engine in Figure 1 (C), Thomas notices that in fact there seems to be an anomaly originating in the B-Bearing, indicated by the dark red colors. This assumption is supported by reviewing the bar chart in (F), where also the B-Bearing is marked as the biggest anomaly in the engine. By hovering over the *Hypermatrix*, he selects orders that can be related to the inner ring of the B-Bearing and the outer ring of the A-Bering as shown in the upper white triangle in (C). He notices that in the line chart and scatterplot the selected order line for the inner ring of the B-Bearing is above a threshold of three, which is marked as red in (F). He can also see that there seems to be a week negative linear relationship between the inner ring of the B-Bearing and the outer ring of the A-Bearing. Thus, he decides to label the engine as B-Bearing error ($\mathbf{T}_5$), which is supported by the fact that all other engines in the cluster are also labeled as B-Bearing error as shown in the engine list view in (B). In the spectrogram in (D) the region where the order of the inner ring of the B-Bearing is shown with a black line and dark grey triangle, he identifies the highest residual values. Therefore, he clicks on "Add Annotation" in (D) and annotates the respective region of the error ($\mathbf{T}_6$). He now confirms his hypothesis that in fact for the sample of 434 engines, B-Bearing errors can also be seen in acoustic measurements of A-Bearings.

8 User Evaluation

In this section, we introduce our evaluation methodology (Section 8.1) and present our study findings (Section 8.2).

8.1 Methodology

Our evaluation has two goals. First, to validate the usefulness and usability of the proposed technical considerations and the resulting visualization in terms of effectiveness for automotive engineers. Second, to evaluate if labeling speed increased after implementing the user feedback from the first evaluation round as part of our iterative design process. The analysis of complex data, where domain knowledge is essential is a high-level cognitive task. Such tasks are however difficult to measure quantitatively and objectively [54]. As for real-world scenarios, data, users, and tasks are important, we perform a qualitative field study to evaluate the usefulness and usability of *IRVINE*. In this type of study, qualitative coding of user feedback is combined with a quantitative usability scale [19]. To evaluate, whether labeling speed increased after including user feedback, we measure how many labels and annotations users made by using the system for twenty minutes. We then compare the results to self-reported labeling speed before the system introduction, before we implemented user feedback, and after the user feedback.

Participants: The study was carried out with six engineers (others than the lead user) from BMW, responsible for the development of testing procedures for the manufacturing of electrical vehicles. They were all male between 23 and 33 years old and had a mean working experience of 4 years in the problem domain, all with a background in mechanical engineering. Inside BMW only very few engineers with a sufficient level of knowledge about the analysis of acoustic measurements of engines exist. Thus, only a low number of potential candidates are able to properly evaluate *IRVINE*. However, this is rather common in design studies, where presented visualizations often tackle very specific problems, which can be addressed by only a few users [11, 19, 47, 48].

Data: For the first round of interviews, we used acoustic data from 434 randomly selected engines over a period of six months. The data did not contain any labels nor annotations. For the second round of interviews, acoustic data from 308 completely new engines from a period of four months were uploaded to the system.

Task: The following task was given to each user: *"Please find error-prone engines and provide labels and annotations for each engine."* Engineers had 20 minutes to find as many error-prone engines as possible. An exemplary execution that we observed during the development with our lead engineer can be the following: First, select a cluster and then an engine from the cluster. Next, analyze the engines' *Hypermatrix*, by hovering over its cells and find a sub-component pair of interest. Then, make a hypothesis, for example, *the resonance of the first gear can also be observed in the rotor-shaft*. Next, inspect the spectrogram, the line charts, the scatterplot, and the bar chart, to accept or reject this hypothesis. If an engine contains an error, select a label from existing labels in the engine list view or create and save a new label as free text and annotate its cause in the spectrogram.

Procedure: As the concept of our *Hypermatrix* and resulting clustering is rather complex, a system introduction was carried out in a kick-off group session. Here, two researchers were present, one tacking notes and the other explaining the system components to the engineers. The session took place online and lasted for 80 minutes. Next, interviews were scheduled with each participant. To evaluate the usability and usefulness of *IRVINE*, four interviews were conducted in the form of a think-aloud session [52]. Each session took on average 60 minutes and involved a short walk-through of the system, open-ended questions [45] about the usage, and a usability questionnaire. All interviews were held online, where each engineer executed the same predefined task. The notes from the kick-off and think-aloud study were analyzed using a qualitative coding methodology [18]. Repeated ideas or statements in the feedback were assigned with codes extracted from the data. Afterward, the codes were grouped into abstract categories to summarize the study results that are also aligned to a set of questions proposed by Lam et al. [33], in the context of user experience. To quantitatively assess the usability of our system, we applied the *System Usability Scale (SUS)* [45]. Due to our small sample size, the SUS scale does not provide empirical evidence of the usability of our visualization but rather a rough direction to support our assumptions of our design choices. In addition, we measured how many labels and annotations were made by the users during the task execution. To evaluate if labeling speed did improve after the implementation of the user feedback, additional two interviews were conducted in the form of a think-aloud session [52]. Here, engineers executed the previously defined task, where we again measured how many labels and annotations were made.

8.2 Findings

We observed that all engineers used the general system workflow as outlined in Section 8.1. First, all engineers selected a cluster, where one engineer also retrained a new SOM and noted, *"I want to have a more detailed view from these similar clusters to better see emerging signatures"*. Next, engineers selected an engine from the engine list. Here, one engineer also did zoom in the cluster, noting that *"for me it is easier, to see all Hypermatrices in one view, to choose a relevant engine"*. All engineers noted that the engine list is very helpful to identify relevant engines. One expert also noted that *"It is good to know that I can filter this list, because I often have specific engine IDs beforehand, which I need to evaluate in detail"*.

Next, engineers turned to the *Hypermatrix* and specifically looked for cross structures in the matrix. This was reported as relevant, where one engineer explained *"This view helps to evaluate how this resonance affects other parts of an engine"*. Next, engineers made hypotheses regarding the signatures of engines and verified them using the spectrogram, the line charts, or the scatterplots. Here, the use of residual values instead of absolute values in decibel was noted to be especially helpful (*"Normally I have to compare two engines, where it is often tedious to find a perfect engine as ground truth."*)

Engineers then assigned labels to the selected engines, where two of them created new label categories in the list view. Three engineers also requested a label from the system after their label input and in all three cases agreed with the systems' suggestion. Next, engineers annotated the cause for the detected error in the spectrogram and reported that it is very helpful that their previous inputs are stored in the system to guide their analyses. One engineer also mentioned that this kind of knowledge helps for discussions with other stakeholders, such as analysts ("It is good to either print it out or use the tool itself to show the analysis to other non-domain related stakeholders").

Engineers also recommended some system improvements, which we implemented in a fourth development cycle in our system. Apart from minor issues, for example, small font sizes, engineers requested to include subcategories for each label. Furthermore, they requested features to better detect relevant engines of interest. This resulted for example in the creation of pie charts that show how many labels exist in each cluster or glyphs, which show the anomaly score of clusters and engines as one can see in Figure 1.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
Expert 1	7.5	10	7.5	5	7.5	7.5	10	10	5	7.5	77.5
Expert 2	7.5	10	7.5	10	7.5	10	5	10	7.5	10	85
Expert 3	10	10	7.5	10	7.5	10	10	7.5	10	10	92.5
Expert 4	7.5	7.5	7.5	7.5	7.5	10	10	7.5	10	10	85
Avg.	8.1	9.4	7.5	8.1	7.5	9.4	8.8	8.8	8.1	9.4	85

Table 1. Results of the System Usability Scale [45] with four engineers.

Considering the quantitative results of the usability survey, our system provides excellent usability according to the adjective equivalent of the achieved SUS score [5]. We found that our systems' usability is with 85 highly above the average score of 68 [45]. The individual scores are outlined in Table 1. Even though we evaluated the system with only four engineers, we were confident that our system did reach a sufficient level of usability and thus focused in later interviews only on evaluating labeling speed. This assumption was supported since the remaining two engineers did not report any recommendations to improve the system.

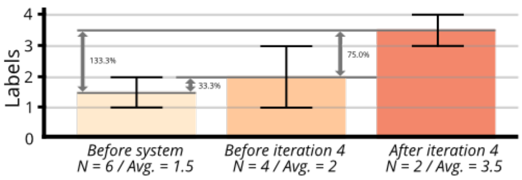


Fig. 10. Improvement of labeling speed compared to the status quo before the introduction of IRVINE, before Iteration 4 and after Iteration 4.

Figure 10 outlines the improvement of labeling speed, which was reached by using *IRVINE*. Before the systems' introduction engineers reported to be able to provide one or two labels in a manual analysis in twenty minutes. However, they noted that it is rather difficult to provide an exact number of labeling speeds since errors and resulting analyses differ significantly. Here, one engineer noted *"There are errors, which can be detected in minutes and others that take up to two hours. However, generally speaking, I normally can provide up to three labels within an hour"*. Even though this number of labels is rather vague, we can use it as a rough direction on how *IRVINE* improved labeling speed.

During our study, we were able to measure the exact number of provided labels during the task execution. As Figure 10 shows, engineers were able to label engines 33.3% faster than before the system introduction and 75.0% after we included the user feedback. If we compare labeling speed to the final system, we can see an improvement of 133%. Even though engineers noted that each analysis heavily depends on the engine and thus labeling speed is generally speaking hard to quantify, we state that with *IRVINE* labeling speed, however, improved.

9 Discussion

In this design study, we contribute the problem characterization and abstraction of the analysis of acoustic signatures in the manufacturing of electrical engines. We further report the interactive design of the presented VA system *IRVINE*, which we evaluated together with six automotive engineers. Our design study represents a very detailed view of the problem domain. From a more abstract point of view, we integrated a VA system into the workflows of domain experts to support them with the interactive clustering and labeling of sensor data. To the best of our knowledge, the problem we described in Section 4.2 is unique inside BMW and has not been addressed by other researchers so far. We thus argue that a design study with a detailed analysis of the problem domain and resulting abstractions was necessary to successfully support domain experts with their very specific tasks.

The overall success of our approach is demonstrated by the fact that study participants rated *IRVINE* with a high usability and usefulness in combination with an increased labeling speed. Our evaluation, however, had a number of limitations. Probably the most important one is the fairly small sample size of participants. However, as mentioned in Section 8.1 it is rather unusual to build solutions for specific domain problems, which address a community bigger than a couple of dozen users. A second limitation is the quantification of labeling speed. As mentioned in our Evaluation, almost every engine has to be treated more or less differently. Especially regarding electrical engines, where product behavior is not that well understood compared to more mature technologies, such as petroleum engines, it is thus hard to determine beforehand how long it will take to produce a label or to carry out a detailed analysis. Nevertheless, we believe that our system increases labeling and analysis speed compared to a manual analysis of the data and serves well to better detect and understand errors in engines.

Although the design was specific to a particular domain, there are some aspects that provide guidance to the design to other domains. Some of that guidance results in the following suggestions to transform high-dimensional data and design similar systems:

1) Cluster Signatures: The SOM visualization helps engineers to immediately detect clusters of similar signatures. The provided small multiples also support the decision, which clusters to analyze first. The detection of a group of potential anomalous engines was known to be an important problem, but before the introduction of *IRVINE*, engineers had to manually compare single engines in a slow error-prone process. In contrast, our SOM visualization allows for a robust and fast identification of relevant engines of interest. The benefits of SOM grid visualizations were previously noted in a use case to analyze speech signals [42].

2) Use Hypermatrices: The *Hypermatrix* view helps experts in the fast allocation of local anomalies in engines. Our approach to compute and visualize *Hypermatrices* can further be used to design systems, where similar data to spectrograms are used as input data. For example, apart from our application domain, spectrograms are applied in the analysis and synthesis of speech signals [23], seismic activities [17], or the medical sector [26], where often sonograms - a similar representation - are applied.

3) Combine Labeling and Annotations: Apart from providing labels for single data instances, as shown in other approaches [44, 58], annotating specific regions directly in matrix views of high-dimensional data can provide more detailed information about the cause of an error. One might argue that suggested labels and annotations from the system are preferable to omit human biases. In our case, however, the primary goal is not only to detect errors but also to understand the reason for their occurrence. This kind of gained knowledge [9] is more relevant than algorithms, which are maybe capable to detect errors fast, but often remain a black box [39]. Thus, we believe that is important to keep experts in an active role in high stakes decision making instead of degrading them to validate or reject model suggestions [2].

4) Make Externalized Knowledge Available: We present guided workflows with our visualization on how to externalize and store knowledge. Here, matrix views, line charts, bar charts, and scatterplots, helped in creating and validating hypotheses. *IRVINE* aggregates knowledge in the form of a label and annotations. This knowledge does not only help domain-related experts to guide their analyses but also other stakeholders outside the application domain [22]. Labels, for example, can be queried by analysts to train machine learning classifiers and annotations can provide valuable insights on the relevant feature space of each label.

We believe that these four recommendations can help other researchers when investigating complex data or domain problems. For instance, Kim et al. [30] and Brattain et al [15] both performed studies on the analysis of sonogram data. To identify secondary symptoms of illnesses they can also compute *Hypermatrices* as demonstrated in our study. Theses *Hypermatrices* could then be clustered by using a similar SOM implementation as demonstrated in Section 5.2. Resulting similar clusters can then be labeled by medical experts and regions of interest directly annotated in the sonograms. This externalized knowledge aids to train other medical experts in the analysis of sonogram data.

The design also had some limitations, which we briefly summarize here. In our visualization, we use data of only one sensor in one test bench. However, using data from different test stations or sensors provides new challenges for visualization designs. One challenge for example is the right mapping of produced components to their recorded data across multiple stations and the visualization of multiple test stations. A solution for this can be the detection of anomalies for multiple stations with network views, and the visualization of according *Hypermatrices* and spectrograms in separate views.

Further, annotations are limited to a rectangular selection. In our case, this is appropriate, since data in the spectrogram has either a vertical or horizontal relation. In use cases where patterns are vertically and horizontally dependent at the same time, annotating data via lasso functionalities might be more adequate.

Finally, to get a comprehensive overview for the *Hypermatrix*, our lead engineer narrowed down 41 relevant orders from the 512 available ones in the spectrograms. This effort was carried out manually and based on years of experience of the engineer in the problem domain. However, this effort can be supported automatically, for example, by computing most deviating orders of a sample of spectrograms to support the selection of important orders before computing the *Hypermatrix*.

10 Conclusion and Future Work

This paper presents a design study on the development of a visualization approach to analyze signatures and acoustic measurements of engines. The resulting VA system *IRVINE* leverages interactive data labeling and clustering approaches to facilitate the analysis of high amounts of acoustic data to detect and understand previously unknown errors. *IRVINE* comprises five different core visualizations. (1) The cluster view and an engine list give an overview over similar signatures of engines. Furthermore, they allow for drilling down to a group of engines and the selection of an engine of interest. *IRVINE* also allows a more fine-grained clustering by retraining the initial SOM based on a selection of sub-clusters and labels. (2) The *Hypermatrix* view allows the analysis of signatures and the creation of hypotheses regarding anomalies in the data. (3) The spectrogram view represents the ground truth of the data to validate previously made hypotheses. (4) Line charts, scatterplots, and bar charts, map *Hypermatrix* selections to spectrograms. (5) Labeling and annotation features allow users to tag specific engines and store knowledge in the system. *IRVINE* evolved iteratively, where we closely worked with engineers from BMW, tested design alternatives, and held critical discussions. The success of our design is shown by the high usability scores, high reported usefulness, and an increased labeling and annotation speed.

There are several avenues for our research. One is to investigate the connection of sensor data across multiple testing stations and the resulting challenges for future visualizations. A second one is to investigate how externalized knowledge enables other stakeholders outside the application domain, for example, data analysts, to improve their work. Finally, the effectiveness of our visualization concept should be investigated in other application domains.

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 1st ed., 2011.

[2] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[3] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In *Visual Analytics Science and Technology*, pp. 3–10, 2009.

[4] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small multipiles: Piling time to explore temporal patterns in dynamic networks. *Computer Graphics Forum*, 34, 2015.

[5] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4:114–123, 2009.

[6] M. Behrisch, F. Korkmaz, L. Shao, and T. Schreck. Feedback-driven interactive exploration of large multidimensional data supported by visual classifier. In *Visual Analytics Science and Technology*, pp. 43–52, 2014.

[7] S. Berg, D. Kutra, T. Kroeger, C. Straehle, B. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Köthe, F. Hamprecht, and A. Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16:1–7, 2019.

[8] J. Bernard, D. Daberkow, D. Fellner, K. Fischer, O. Koepler, J. Kohlhammer, M. Runnwerth, T. Ruppert, T. Schreck, and I. Sens. Visinfo: a digital library system for time series research data based on exploratory search—a user-centered design approach. *International Journal on Digital Libraries*, 16(1), 2015.

[9] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *Transactions on Visualization and Computer Graphics*, PP:1–1, 2017.

[10] J. Bernard, C. Ritter, D. Sessler, M. Zeppelzauer, J. Kohlhammer, and D. Fellner. Visual-interactive similarity search for complex objects by example of soccer player analysis. In *Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 3, pp. 75–87, 2017.

[11] J. Bernard, D. Sessler, J. Kohlhammer, and R. A. Ruddle. Using dashboard networks to visualize multiple patient histories: A design study on post-operative prostate cancer. *Transactions on Visualization and Computer Graphics*, 25(3):1615–1628, 2019.

[12] J. Bernard, N. Wilhelm, B. Krüger, T. May, T. Schreck, and J. Kohlhammer. Motionexplorer: Exploratory search in human motion capture data based on hierarchical aggregation. *Transactions on Visualization and Computer Graphics*, 19(12):2257–2266, 2013.

[13] J. Bernard, M. Zeppelzauer, M. Sedlmair, and W. Aigner. Vial: a unified process for visual interactive labeling. *The Visual Computer*, 34, 2018.

[14] M. Bostock, V. Ogievetsky, and J. Heer. D-3: Data-driven documents. *transactions on visualization and computer graphics*, 17:2301–9, 2011.

[15] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, and A. E. Samir. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdominal Radiology*, 43(4):786–799, 2018.

[16] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. Dis-function: Learning distance functions interactively. In *Visual Analytics Science and Technology*, pp. 83–92, 2012.

[17] V. W. Chao, Y. Wu, L. Zhao, V. Tsai, and C.-H. Chen. Seismologically determined bedload flux during the typhoon season. *Scientific Reports*, 5, 2015.

[18] K. Charmaz. *Constructing grounded theory : a practical guide through qualitative analysis*. Sage Publications, London; Thousand Oaks, Calif., 2006.

[19] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. Paved: Pareto front visualization for engineering design. *Computer Graphics Forum*, 39:405–416, 2020.

[20] J. J. Dudley and P. O. Kristensson. A review of user interface design for interactive machine learning. *Transactions on Interactive Intelligent Systems*, 8(2), 2018.

[21] J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *Visualization in Data Science at IEEE VIS*, 2020.

[22] J. Eirich, D. Jäckle, S. Werrlich, and T. Schreck. Visual analytics in organizational knowledge creation: A case study. In *European Conference on Information Systems*, 04 2021.

[23] J. L. Flanagan. *Speech analysis; synthesis and perception*. Springer, 2nd ed., 1972.

[24] M. Hardin, D. Hom, and L. Williams. Which chart or graph is right for you?, 2016.

[25] F. Heimerl, S. Koch, H. Bosch, and T. Ertl. Visual classifier training for text document retrieval. *Transactions on Visualization and Computer Graphics*, 18(12):2839–2848, 2012.

[26] S. Hughes. Medical ultrasound imaging. *Physics Education*, 36:468, 2001.

[27] B. Höferlin, R. Netzel, M. Höferlin, D. Weiskopf, and G. Heidemann. Inter-active learning of ad-hoc classifiers for video visual analytics. In *Visual Analytics Science and Technology*, pp. 23–32, 2012.

[28] A. Kampker, K. Kreiskother, N. Lutz, V. Gauckler, and M. Hehl. Re-ramp-up management of scalable production systems in the automotive industry. In *Industrial Technology and Management*, pp. 137–141, 2019.

[29] D. A. Keim. Information visualization and visual data mining. *Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[30] S. Kim, B. K. Seo, J. Lee, K. Cho, K. Lee, B.-K. Je, H. Y. Kim, Y.-S. Kim, and J.-H. Lee. Correlation of ultrasound findings with histology, tumor grade, and biological markers in breast cancer. *Acta oncologica*, 47:1531–8, 2008.

[31] T. Kohonen. Essentials of the self-organizing map. *Neural Networks*, 37:52–65, 2013.

[32] T. Kohonen, M. R. Schroeder, and T. S. Huang. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 3rd ed., 2001.

[33] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *Transactions on visualization and computer graphics*, 18, 2011.

[34] J. Moehrmann, S. Bernstein, T. Schlegel, G. Werner, and G. Heidemann. Improving the usability of hierarchical representations for interactively labeling large image data sets. In *Human computer interactions*, p. 618–627. Springer, Berlin, Heidelberg, 2011.

[35] T. Munzner. A nested model for visualization design and validation. *Visualization and Computer Graphics*, 15:921 – 928, 2010.

[36] I. Nonaka and H. Takeuchi. *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford University Press, New York, 1995.

[37] A. Pister, P. Buono, J.-D. Fekete, C. Plaisant, and P. Valdivia. Integrating prior knowledge in mixed initiative social network clustering, 2020.

[38] C. Ratanamahatana, J. Lin, D. Gunopulos, E. Keogh, M. Vlachos, and G. Das. *Mining Time Series Data*, pp. 1049–1077. Springer, 2010.

[39] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215, 2019.

[40] T. Ruppert, M. Staab, A. Bannach, H. Lücke-Tieke, J. Bernard, A. Kuijper, and J. Kohlhammer. Visual interactive creation and validation of text clustering workflows to explore document collections. *Electronic Imaging*, 2017(1):46–57, 2017.

[41] D. Sacha, Y. Asano, C. Rohrdantz, F. Hamborg, D. Keim, B. Braun, and M. Butt. Self organizing maps for the visual analysis of pitch contours. *Computational Linguistics*, 2015.

[42] D. Sacha, M. Kraus, J. Bernard, M. Behrisch, T. Schreck, Y. Asano, and D. A. Keim. Somflow: Guided exploratory cluster analysis with self-organizing maps and analytic provenance. *Transactions on Visualization and Computer Graphics*, 24(1):120–130, 2018.

[43] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. Visual interaction with dimensionality reduction: A structured literature analysis. *Transactions on Visualization and Computer Graphics*, 23(1):241–250, 2017.

[44] A. Sarkar, M. Spott, A. Blackwell, and M. Jamnik. Visual discovery and model-driven explanation of time series patterns. In *Visual Languages and Human-Centric Computing*, pp. 78–86, 2016.

[45] J. Sauro. Measuring usability with the system usability scale (sus), Accessed 05.11.2020.

[46] T. Schreck, J. Bernard, T. von Landesberger, and J. Kohlhammer. Visual cluster analysis of trajectory data with interactive kohonen maps. *Information Visualization, Palgrave Macmillan*, 8(1):14–29, 2009.

[47] M. Sedlmair, A. Frank, T. Munzner, and A. Butz. Relex: Visualization for actively changing overlay network specifications. *Visualization and Computer Graphics*, 18:2729–2738, 2012.

[48] M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz. Cardiogram: Visual analytics for automotive engineers. In *Human Factors in Computing Systems*, pp. 1727–1736, 2011.

[49] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics*, 18:2431–2440, 2012.

[50] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.

[51] B. Shneiderman. Inventing discovery tools: Combining information visualization with data mining. *Information Visualization*, 1(1):5–12, 2002.

[52] M. Someren, Y. Barnard, and J. Sandberg. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses*. Academic Press, 1994.

[53] J. Suschnigg, B. Mutlu, A. Fuchs, V. Sabol, S. Thalmann, and T. Schreck. Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Big Data Visual Exploration and Analytics*, 2020.

[54] M. Tory and T. Möller. Evaluating visualizations: Do expert reviews work? *Computer graphics and applications*, 25:8–11, 2005.

[55] J. Vesanto. Som-based data visualization methods. *Intelligent Data Analysis*, 3(2):111–126, 1999.

[56] N. Weber, M. Waechter, S. C. Amend, S. Guthe, and M. Goesele. Rapid, detail-preserving image downscaling. *Transactions on Graphics*, 6, 2016.

[57] N. Wilhelm, A. Vögele, R. Zsoldos, T. Licka, B. Krüger, and J. Bernard. Furyexplorer: Visual-interactive exploration of horse motion capture data. *Visualization and Data Analysis*, 2015.

[58] D. Wu, X. Wang, J. Su, B. Tang, and S. Wu. A labeling method for financial time series prediction based on trends. *Entropy*, 22(10), 2020.

[59] W. Yang, X. Wang, J. Lu, W. Dou, and S. Liu. Interactive steering of hierarchical clustering. *Transactions on Visualization and Computer Graphics*, 2020.

[60] J. Zhao, M. Karimzadeh, A. Masjedi, T. Wang, X. Zhang, M. M. Crawford, and D. S. Ebert. Featureexplorer: Interactive feature selection and exploration of regression models for hyperspectral images. In *Visualization Conference*, pp. 161–165, 2019.

6.4 P4: ManEx: The Visual Analysis of Measurements for the Assessment of Errors in Electrical Engines

Full reference: *J. Eirich, G. Koutroulis, B. Mutlu, D. Jäckle, R. Kern, T. Schreck, and J. Bernard. Manex: The visual analysis of measurements for the assessment of errors in electrical engines. IEEE Computer Graphics and Applications, pp. 1–13, 03 2022*

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

Department: Head
Editor: Name, xxxx@email

ManEx: The Visual Analysis of Measurements for the Assessment of Errors in Electrical Engines

J. Eirich
University of Bamberg, BMW Group

G. Koutroulis
Pro2Future GmbH

B. Mutlu
Pro2Future GmbH

D. Jäckle
BMW Group

R. Kern
Graz University of Technology

T. Schreck
Graz University of Technology

J. Bernard
University of Zurich

***Abstract*—Electrical engines are a key technology all automotive manufacturers must master to stay competitive. Engineers need to analyze an overwhelming number of engines measurements to improve the manufacturing for this technology. They are hindered in the task of analyzing large numbers of engines, however, by the following challenges: (1) Engines comprise a complex hierarchical structure of subcomponents. (2) Locating error causes along the manufacturing processes is a difficult procedure. (3) Large numbers of heterogeneous measurements impair the ability to explain errors in engines. We address these challenges in a design study with automotive engineers and by developing the Visual Analytics system *Manufacturing Explorer* (ManEx), which provides interactive interfaces to analyze measurements of engines across the manufacturing process. ManEx was validated with five experts. Our results suggest high usability and usefulness scores and the improvement of a real-world manufacturing process. Specifically, with ManEx experts reduced scraped parts by over 3%.**

Introduction

E-mobility represents one of the largest transformations for the automotive industry in the last 100 years [1]. In this context, electrical engines (hereafter: engines) are one of the key technologies companies need to master to stay competitive. To meet high-quality standards, manufacturers are leveraging manufacturing data for improving production processes. In this regard, engineering experts must analyze enormous numbers of sensor measurements of thousands of engines across the manufacturing process. This results in large data quantities, which must be analyzed by human experts. However, engineers are hindered in their efforts to effectively analyze engines by the following three challenges:

(1) Hierarchical challenge: A single engine is the result of the assembly of many complex parts. The composition of the parts forms a complex hierarchical structure. If there is an error in an engine, identifying whether this can be related to individual parts or to the composition of interrelated parts is a very difficult task. Engineers thus need an efficient means of relating errors either to individual parts or to the composition of interrelated parts. This is important for ensuring that the manufacturing process can proceed on time and for assuring that high-quality engines can be delivered.

(2) Temporal challenge: An engine is produced in many sequential steps. To increase manufacturing output, each step is executed by multiple identical stations. Furthermore, engineers aim to automatically identify errors at designated test stations. These test stations are often located in later manufacturing steps, which results in the problem that errors are also detected late. Finding out how the error cause can be identified in the early steps is thus another challenge.

(3) Measurement challenge: During the manufacturing process, a great number of measurements are recorded at each station. Since all the stations record different measurements unifying them in a single approach and understanding their relations is a major challenge.

To address these challenges, we worked together in a *design study* with BMW engineers with extensive knowledge and experience in the manufacturing of engines [2].

To help engineers with their tasks of analyzing large numbers of engines and measurements, we first familiarized ourselves with their domain, gathered data across the manufacturing process, and gained an understanding of their problems. We then sub-divided the domain problem into an approach, which visualizes the hierarchical and temporal structure of the manufacturing process and provides help in analyzing measurements by integrating *anomaly detection* and *causal discovery*. As a result of this collaboration, we present the Visual Analytics (VA) system ManEx (*Manufacturing Explorer*).

ManEx provides interactive interfaces to analyze measurements of engines across the manufacturing process. Engineers can use their domain knowledge to analyze anomalies in measurements, and judge whether an anomaly serves identify engines, which might either include errors. Furthermore, causal discovery provides insights to find the root cause of detected errors. For a detailed analysis, measurements can be added to a subset of measurements, which we refer to as *focus-set*. Engineers can evaluate the quality of a *focus-set* until it is capable of detecting errors in the manufacturing process. In summary, we contribute:

(1) The problem characterization and abstraction of the domain for analyzing large numbers of measurements and engines in manufacturing processes.

(2) *ManEx*: An interactive VA system for the analysis of large-scale manufacturing data.

(3) An evaluation of *ManEx* with five automotive engineers, as well as reflections of our design process.

Related Work

First, we introduce VA applications in the automotive sector, followed by use cases of anomaly detection in the manufacturing industry and insights on causality-aided VA.

Automotive Visual Analytics Applications

Visualization to date has contributed substantially to help analyzing complex data in industrial applications, as the recent survey by Zhou et al. [3] shows. A key important domain is the automotive sector. There, VA applications can support the engineering design, the optimization of testing procedures, condition monitoring of assembly lines, or the visual support of high cognition tasks. Important efforts were focused, among others on visualizing in-car communication networks [4], multi-criteria alternatives for the exploration of rotor designs [5], the visual exploration of assembling data to detect inefficiencies [6], or to support domain experts in the analysis of acoustic signatures [1]. All studies specifically address high cognition tasks, which can be performed by experts in organizations. However, these studies focus on data from one source (e.g., single station). Grounded on previous findings, we build a system that visualizes data from multimodal sources across an entire manufacturing process, supporting domain experts in their high cognition task of analyzing measurements of engines.

Anomaly Detection in Manufacturing

Potential errors in measurements can be supported using anomaly detection. So far, a wide range of anomaly detection with VA has been introduced in the manufacturing sector. Wörner et al. [7] visualize diagnostic machine data to help identify specific elements in machines, which need to be replaced or repaired. Maier et al. [8] present a visualization to guide the user in detecting anomalies of time series data in model manufacturing plants. Eirich et al. [9] computed anomalies from time series manufacturing data for engines and developed a system, which visualizes the most relevant anomalies. Furthermore, Suschnigg et al. [10] consider anomaly detection in multivariate times series from test benches and display anomalies as a collection of colorencoded cycle glyphs. All studies show that visualizing anomalies of manufacturing data helps engineers to better identify errors in the manufacturing process. Thus, we build on previous findings of anomaly detection and visualization. Particularly inspiring for the visualization of anomalies is the representation of anomalies with cycle glyphs [10].

Causality-aided Visual Analytics

A critical question in manufacturing processes is "What are the causes of significant changes in downstream processes?". Such questions can be answered with causal discovery, where the underlying causal information from observational data is inferred as directed graphs. The causal structure of a manufacturing process, however, is often too complex for users to understand by only looking at the outcome of algorithms. Thus, a commonly used tool to display the causal relations between parameters is a directed graph [11]. Regarding the automotive sector, Dentzner et al. [12] propose a VA system for root cause analyses in an assembly line. The authors employed interactive decision trees, which in our case are misleading since the underlying causal information is extracted exclusively via correlational techniques. In turn, we enhance *ManEx* with the PC algorithm [13], a constraint-based causal discovery method, as it exploits sparsity in high-dimensional datasets, which are ubiquitous in manufacturing environments.

Methodology and Domain Problem

At the beginning of the project, we worked together with three automotive engineers of BMW and learned about their problems, goals, and tasks when analyzing manufacturing data. The exchange with them took place from January until March 2021 with an average of two meetings a week ranging from 20-90 minutes. After we developed an understanding of the problems, goals, and tasks, we began to design and develop *ManEx* in six main iterations for six months. During each iteration, we interviewed engineers, tested design alternatives, and held constructive discussions with visualizing experts. An evaluation of the system was carried out with five engineers at BMW (others than the engineers we developed the system with).

The Domain of Manufacturing Engines

To demonstrate the assembly of an engine as it is built at BMW, we provide an example in Figure 1. An engine consists of four parts: rotor (A), stator (B), gear (C), and inverter (D). A rotor and a stator together form an electrical machine (A+B), which we hereafter refer to as EMA. Finally, an EMA is assembled with the gear and inverter to form an engine.

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

Figure 1: Topology of an engine at BMW. An engine consists of the parts A (rotor), B (stator), C (gear), and D (inverter).

The engineering experts reported that the assembly of an engine can be divided into a *temporal* and *hierarchical* dimension. Both dimensions can also be visualized as flow diagram as shown in Figure 2 and be described as follows:

Temporal dimension: An engine is assembled along its temporal dimension. For example, a stator and a rotor are together assembled into an EMA. After that, the EMA, inverter, and gear are assembled into an engine.

Hierarchical dimension: Each part of an engine inherits manufacturing steps, which again are hierarchically structured into stations. In this regard, steps have a temporal relation meaning that each step is executed sequentially. In turn, stations are all identical and run in parallel to increase manufacturing output. At the root of each station are measurements.

Domain Problem

Following the information received from the engineers, we can define the domain problem for the *hierarchical, temporal* and *measurement* challenge as follows:

Where in the hierarchical dimension? Engineers analyze data only for individual parts, steps, and stations. Connecting data from multiple sources is performed manually. As a result it is difficult to determine whether an encountered error can be related to one part, step, or station, or if it is the result of the composition of interrelated parts, steps, or stations. The experts thus cannot identify the location of an error simply by means of efficiently analyzing data from multiple parts, steps, or stations.

When in the temporal dimension? Currently, engineers manually combine measurements from engines. They reported, however, that they find it difficult to identify when the error first appeared in the manufacturing process. In addition to the problem of connecting data from parts, steps, and stations with each other, statistical analyses to infer relations along the temporal dimension must be programmed manually by engineers. Even though it is important to identify the root cause of errors as fast as possible, engineers have no system at hand to support them with the analysis of the temporal dimension.

How in the measurements? Furthermore, engineers have to analyze large numbers of heterogeneous measurements, which is simply not feasible for a detailed manual analysis. Even though it is possible to connect data from various sources and to manually perform statistical analyses, engineers reported that it is difficult to know how measurements influence errors. It was noticed by the engineers, however, that the analysis of anomalies is a promising direction for gaining a better understanding of the relationship between measurements. In accordance with the engineers, we consider anomalies as irregularities in measurements including statistical outliers. The severity of an anomaly is left to the judgment of an engineer.

Abstractions

We abstract from the domain of the engineers, the domain problem, and the terminology. The resulting characterization of the analyzed data and tasks are presented in the following subsections.

Data Characterization

Along the *hierarchical* and *temporal* dimensions, engineers analyze measurements from engines to identify and understand errors. Both data types can be abstracted as follows:

Engines: Engineers analyze several thousand engines and never only one. Thus, engines represent the analysis object and can be divided into two analysis targets:

(1) *Engines with explainable error class* is a group of engines with a specific error, which experts already know before performing an analysis. In this analysis, the goal lies for example in identifying the root cause of an error. For example, a

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

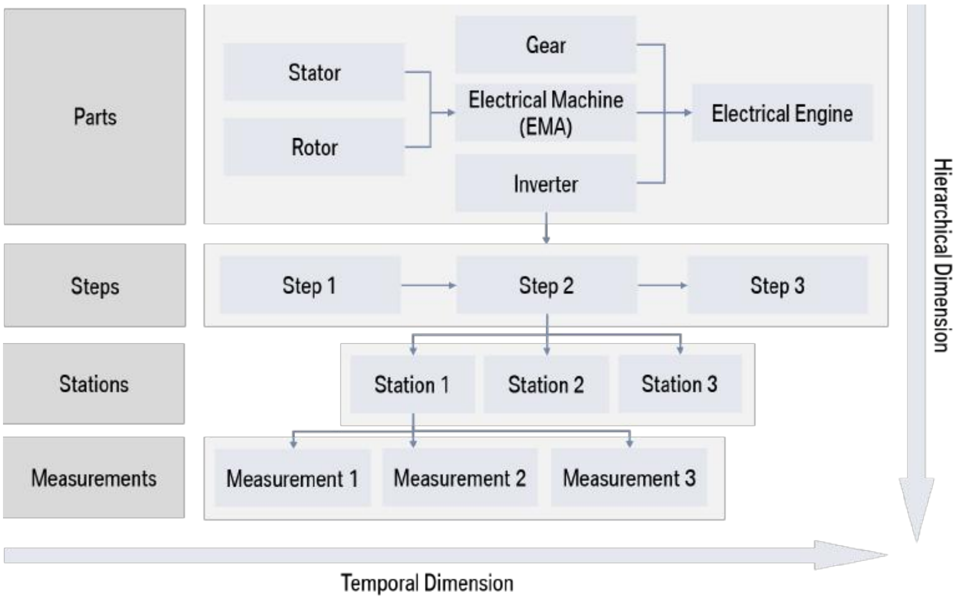


Figure 2: An engine is assembled along two dimensions. On the temporal dimension, the assembly and step sequences are represented. The hierarchical dimension represents the hierarchy of parts, steps, stations, and measurements.

rotor can have a bad electrical conductivity, which is detected at an electrical test bench for an EMA. However, the cause of this error might lie in a welding step of copper wires of the rotor.

(2) *Engines with novel unexplainable error class* is a group of engines that does not have a specific error before an analysis begins and where the goal is about a general analysis of engines to identify and understand new errors. For example, multiple engines can sound loud, when measured in an acoustic test bench. However, this phenomenon might be new to engineers and could be either a new error type or simply a statistical outlier.

Measurements: Each station records heterogeneous measurements (pressure or voltage, among others). More than 10,000 measurements are recorded for an engine. Measurements cannot be compared with each other due to their heterogeneity (e.g., voltages vs. pressure). The engineers suggested using *residual values* for each measurement instead of absolute values. As a result, the deviation from expected measurement values will stand out, which makes the explanation of identified errors easier. In addition to engines, measurements also can be divided into the following two categories:

(1) *Known measurements* are measurements experts already know before an analysis. For example, engineers are generally responsible for a specific test station. If an anomaly is measured inside their station, engineers already have multiple hypotheses at hand on how this anomaly can be interpreted. For example, engineers know exactly which regions of an engine's acoustic signature can be related to a rotor or a stator.

(2) *Unknown measurements* are measurements experts do not know before an analysis begins. This analysis is about identifying new measurements, which might be relevant for an error. For example, engineers might not be aware that an electrical signature of an engine can be highly significant for its acoustic signature.

Task Abstraction

In the following, we present the task abstraction. The presented tasks (**T**) will serve as primary design targets for *ManEx*. We base the tasks on the task taxonomy of Brehmer and Munzner [14]. This taxonomy is appropriate because it forms the bridge between high and low-level tasks, which is particularly helpful for embedding them into the daily routines of the engineers. Our resulting tasks are displayed in Figure 3, which can be

	Engines with explainable error class	Engines with novel unexplainable error class	Additional comparison and information seeking tasks
Measurements known	**Task 1**: Lookup engine group and measurements	**Task 3**: Browse for engine group	**Task 5**: Compare engine groups
Measurements unknown	**Task 2**: Locate measurement of interest	**Task 4**: Explore engines and measurements	**Task 6**: Identify relation between measurements

Figure 3: Derived tasks for ManEx following the engineers from BMW and the task taxonomy of [14]. Engines can either contain explainable error classes or new unexplainable error classes. To identify and understand errors, engineers can use measurements they already know, or unknown measurements. Along the two dimensions, engineers perform the four tasks T1-4. While performing these tasks, engineers compare engine groups (T5) or identify relations between measurements (T6).

summarized as follows:

T1 *Lookup engine group and measurement*: Engineers need to lookup known measurements for an engine group with an explainable error. Thus, the system should support the fast and efficient loading of engines with an error into the system and an easy navigation to the known measurements.

T2 *Locate measurement of interest*: Engineers, who want to identify new relevant measurements for an engine with an explainable error need a drill-down mechanism to locate measurements of interest. The information requirement may differ between the severity of anomalies for groups of engines with an error, or causal relationships between measurements.

T3 *Browse for engine group*: When measurements are known and the error for engines is not yet explainable, *ManEx* needs to support the browsing for a group of interesting engines. These engines may contain unnoticed errors, or an unusually high amount of anomalies, which might be harbingers of future errors. The information need may differ to the severity of anomalies of measurements in parts, steps, and stations.

T4 *Explore engines and measurements*: This task is about getting a general overview of the hierarchical and temporal dimensions of the manufacturing process. Thus, interactive means that allow rapid and easy navigation to measurements of interests, or to select an engine group for an in-depth analysis, should be supported.

T5 *Compare engine groups*: While performing the tasks T1-4, engineers need support in comparing groups of engines with each other.

T6 *Identify relations between measurements*: As well as **T5**, engineers also perform this task while executing T1-4. Here, *ManEx* has to support engineers in identifying relations between measurements.

Data Preprocessing

In this section, we introduce the computation of anomalies, followed by the identification of causal relations between measurements. We then proceed to illustrate technical details on the evaluation of selected measurements.

Anomaly Computation

We provide the following two methods to detect anomalies in measurements: *anomaly scores*, which focus on measurements (1), and *t-tests*, which focus on engines (2).

In (1), for each measurement m_i, we calculate the mean and standard deviation for an ensemble of engines. Consequently, the relative deviation of the i-th measurement to the mean μ for each is then expressed in units of the standard deviation σ as outlined in Equation 1:

$$\text{Residual}_i = \frac{\text{m}_i - \mu}{\sigma}. \tag{1}$$

Next, we aggregate the residual values for all measurements in a station as outlined in Figure 4. We calculate anomaly scores for each station by computing the mean, median and the max/min of the aggregated residual values. We repeat this process for all stations and steps up to a part level.

We also experimented with different approaches to compute anomalies. Instead of com-

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

puting residual values, we calculated the deviation of each absolute measurement value to fixed thresholds, which are used in the manufacturing process. However, engineers reported that what they want is to have a measure, which is not based on a static threshold but depends on the distribution of the measurements in order to get a new perspective on the data they analyze.

Furthermore, measurement recordings might slightly change over time (e.g., different temperatures in winter and summer in a manufacturing facility). Thus, residual values are preferable, since they consider the deviation of a single measurement recording to a large sample of measurement recordings instead of relying on a static threshold. In (2), we use a statistical measurement that indicates whether or not two groups of engines are different from each other. Since the following conditions can be applied to our data [15], two-sided t-tests are an appropriate choice:

1) The goal of our measurements is to derive a hypothesis about the difference of an error in one engine group to those of all the other engines.

2) The independent variable we analyze is quantitative.

3) We use two samples, which are engines with an error and engines with no error.

4) We look at dependent distributions, this being the entire population against a group of engines with an error out of the same population.

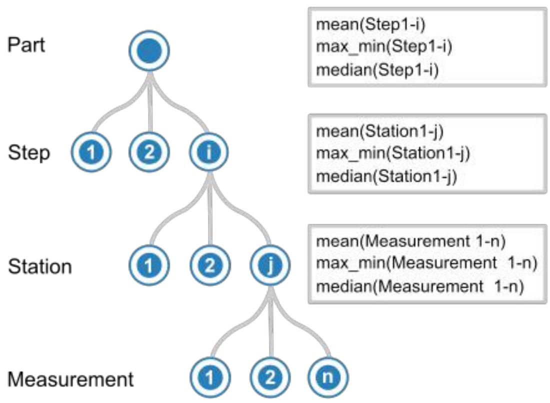


Figure 4: Aggregation of anomalies on a part, step and station level based on measurements. An anomaly propagates upwards in the hierarchy. Users can choose between the aggregation methods: mean, max/min, and median.

Causality Computation

To infer the causal graph from all investigated measurements, the PC algorithm [13] is employed. This method utilizes conditional independence testing to first eliminate causal edges from a fully connected undirected graph, and second to propagate all orientations at the remaining edges. We integrate prior knowledge to the PC algorithm from the temporal ordering of the stations and steps for each part. Note that in each station no such ordering constraint is available, since they all run in parallel. For example, based on the arrangement of the manufacturing stations for each engine (see Figure 2), stations for both stator and rotor precede EMA. Formally, $\mathbf{V}^s$ the set of measurements from the stator, $\mathbf{V}^r$ the set of measurements from the rotor and finally $\mathbf{V}^{ema}$ the set of measurements from EMA. Since dependence of upstream stations is infeasible based on our assumption, we eliminate all edges $\mathbf{V}^{ema} \rightarrow \mathbf{V}^s$ and $\mathbf{V}^{ema} \rightarrow \mathbf{V}^r$. Furthermore, the tasks at the aforementioned stations for stator and rotor may be performed concurrently, thus rendering them jointly independent. Therefore, we additionally eliminate all edges $\mathbf{V}^s \rightarrow \mathbf{V}^r$ and $\mathbf{V}^r \rightarrow \mathbf{V}^s$. Finally, we obtain a sparse causal graph by incorporating valuable domain knowledge that supports the identification of the root cause of the concerned measurements. The domain knowledge is about the hierarchical and temporal dimensions of the manufacturing process, as well as the measurements. Examples of such domain knowledge are how different parts are assembled to an engine, or which manufacturing step, or which measurement is recorded.

The PC algorithm is comprised mainly of two free parameters, the significance level α and the choice of the conditional independence test for evaluating the null hypothesis of independence. We set a relatively conservative value ($a = 0.01$) since higher values will lead to more dense causal graphs and ultimately high uncertainty due to the increased number of false positives. Finally, we employ Fisher's z-transformation of the partial correlation as the conditional independence test, as we already assumed that all measurements are normally distributed.

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

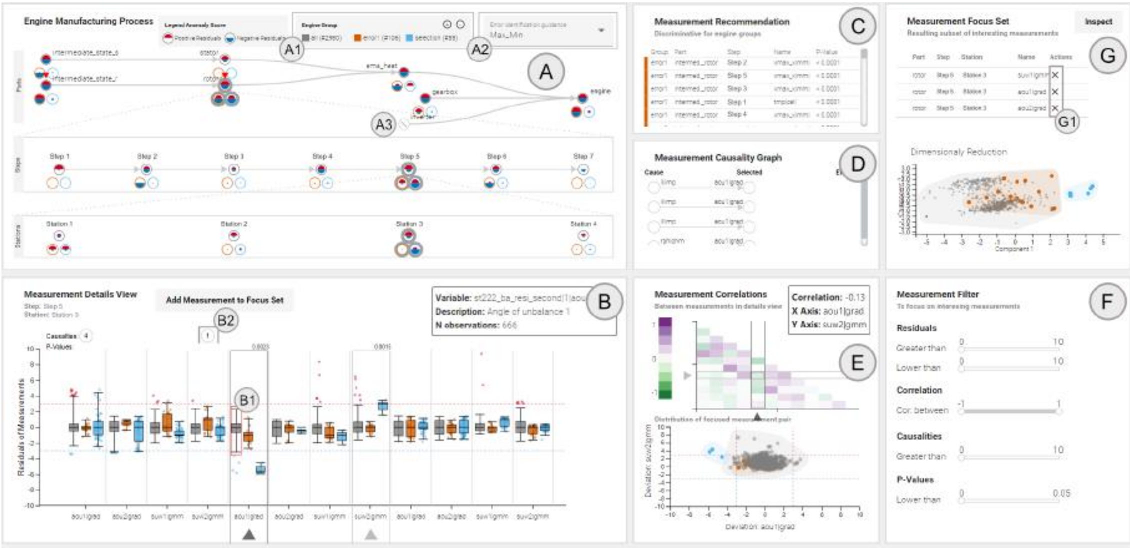


Figure 5: The ManEx System: Users have an overview of parts, steps, and stations in (A). Measurement details for selected stations are shown in (B). Users can locate relevant measurements based on t-tests in (C) or causalities in (D). The correlation for a selected sample of measurements is shown in (E). A sample of measurements can be filtered in (F) and the strength of the computed relationships be evaluated in (G).

Focus-Set Evaluation

The engineers reported that they wish to sequentially add measurements to the *focus-set* to identify the relation between measurements. When more than three measurements are selected for an analysis, these cannot be visually represented. Hence, applying dimensional reduction to the *focus-set* is an adequate choice. We use principal component analysis (PCA), which is a well-established method for identifying linear relationships in high-dimensional data. We could also use methods for discovering non-linear relationships, such as t-Distributed Stochastic Neighbor Embedding (*TSNE*). However, *TSNE* requires a manual parameterization, which engineers would need to carry out each time they analyze a *focus-set*. Hence, we consider PCA as the most straightforward method to evaluate the quality of the *focus-set*.

Visualization

Figure 5 shows an overview of the views of *ManEx* marked from A-F. Due to non-disclosure agreements with BMW, steps are named step 1-n and stations respectively station 1-x. All color schemes were selected in close collaboration with the engineers and by using colormaps from the open-source *ColorBrewer* system [16]. We paid close attention to ensuring that no color was used twice, and that all color schemes account for color deficiencies.

Overview over Hierarchical and Temporal Dimensions

In (A), the hierarchical and temporal dimensions are displayed. Selecting a part unfolds its steps, selecting a single step further allows users to drill down to stations involved in this step (**T1**). Selections at all three levels of the hierarchy are always outlined with a thick grey stroke, dashed gray lines indicate the unfolded hierarchy. All engine groups are displayed in the legend control (A1); qualitative colors support the differentiation between engine groups across views (A), (B), (C), (E), and (G); the (control) group with all engines is always shown with gray color (**T5**). Forming new groups is enabled in two ways: Users can either use pre-defined groups using CSV file upload (**T1**) in (A1) – the observed default behavior among engineers – or manually add a group of engines selected in associated views (**T3**) (B) or (G). In (A2), users can choose between the mean, max/min, or median as a measure for anomaly scores as shown in Figure

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

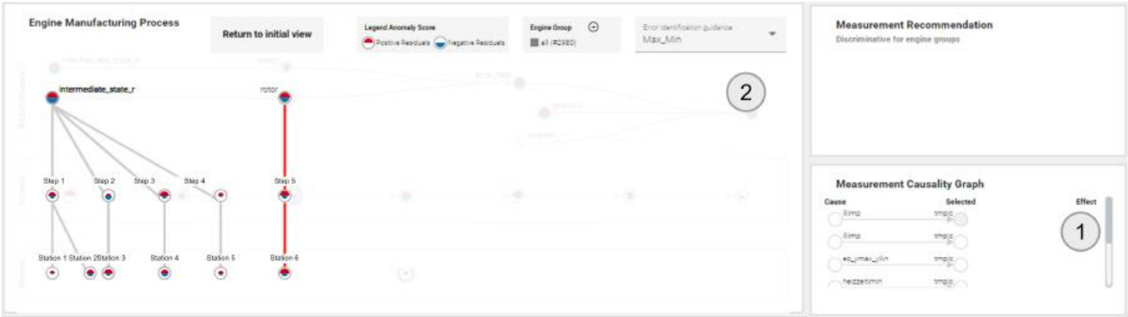


Figure 6: Causality graph for a measurement. Measurements in the graph are displayed in (1). Hovering over nodes in (1) results in an overview of where along the hierarchical and temporal dimension of the manufacturing process the causalities are located in (2).

4. Parts that do not contain data are shown with a gray line as shown in (A3). A legend showing positive values for the anomaly scores in red and negative ones in blue is displayed left to (A1).

Users have two options for drilling down to measurements of interest (**T2**). First, they can manually navigate through parts, steps, and stations evaluating the respective anomaly scores and their relations across the hierarchical and temporal dimensions of the manufacturing process (**T6**). Here, anomaly scores are displayed as two colored glyphs, where the arc size indicates the severity of an anomaly. Second, they can select a measurement using t-tests, which are shown in (C) with a table view and ordered by their p-value. By either selecting a station from (A) or a row from (C), the according measurements are shown in (B).

Measurement-Centric Views

The distribution of all measurements for a selected station is displayed in (B). We use boxplots to ease measurement comparison side-by-side. Measurements are aligned along the x-Axis and separated with vertical gray lines. In addition, users can also compare the distributions for individual engine groups (A2) within each measurement (**T5**). According to the engineers, it is always interesting to identify measurements for which different engine groups behave in considerably different ways (**T4**). We support this task with automatically calculated t-tests for every measurement, to seek p-values supporting assumptions of differences across engine groups. If available, these p-values are displayed above each measurement (B1).

Users can further assess the relations across any pair of measurements (**T6**) in (E) with a correlation matrix, with a diverging green-to-purple color scale to encode Pearson correlations between -1.0 and 1.0.

A mapping from the selected correlation pair in (E) to (B) is performed with gray triangles in (E) and (B) and lines in (B). A tooltip displays more information for the selected correlation pair in (E). Furthermore, hovering over the correlation matrix also shows the detailed distribution of the two selected measurements as a scatterplot below the matrix. Users can apply filters in (F) to reduce the number of visible measurements in (B), especially if stations of interest contain many measurements. Also, users can add interesting measurements to the *focus-set* by clicking on a boxplot as shown in B1 with a red outline.

Causal Discovery

When users aim at identifying relevant measurements outside a station analyzed in (B), relation-seeking and causal discovery are needed (**T6**). The available causalities as well as p-values appear above each measurement. On selecting a causality in (B2) its causal graph is shown in (D), and relevant parts are marked with a red triangle in (A). A selected measurement can either be influenced by measurements of previous steps (Cause) or have an impact on the measurements of later steps (Effect). By hovering over nodes in (D), only the relevant parts, steps, and stations are shown as a second layer above (A) as demonstrated in Figure 6. The selected node

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

from Figure 6 (1) is shown with a red outline in Figure 6 (2). Clicking on a node freezes the layer above (A) and replaces all boxplots in (B) with the measurements of the causality set. Of course, users can always return to the initial view.

Measurement Validation

Measurement relations (**T6**) can be analyzed in the *focus-set* in (G). All selected measurements are displayed in a table view. As soon as more than two measurements are in the *focus-set*, the PCA is executed and the result displayed as a scatterplot, using a similar visual metaphor as used in (E) (**T4**). Measurements can be deleted in (G1). After measurements are added or deleted, the PCA changes accordingly. By clicking on the button "Inspect", the selected measurements replace the boxplots in (B) and correlations in (E).

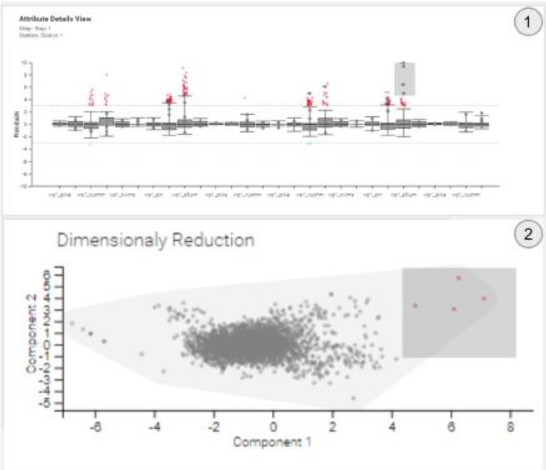


Figure 7: A brush Selection for engines of interest in boxplots (1) and the PCA scatterplots in (2), where the selection in (1) is highlighted with additional gray circles in the other boxplots.

Manual Subset Selection

Users can manually browse engines for in-depth analyses (**T3**). To do so, they perform a brush selection as demonstrated in Figure 7 (1) for the boxplots (B) and Figure 7 (2) for the scatterplot (G). In the case of the boxplots in Figure 7 (1), the values of the same selected engines in the station are highlighted with gray circles. After the selection, *ManEx* performs the same operations as when manually uploading an engine group in (A2).

Iterative Design Process

The design of *ManEx* was carried out in six main iterations in close collaboration with the engineers. After each iteration, we gathered feedback and adapted our visualization design accordingly. The six iterations (**I1-6**) can be briefly summarized as follows:

I1 *System back-end:* The first iteration was mainly about setting up the system back-end. This included the mapping of engine measurements to the steps and stations. This iteration was probably the most time intensive step in the entire project. Nevertheless, no visualization design was carried out here.

I2 *Basic system functionalities:* In a first attempt, we used a parallel coordinates view. However, this view implied a linear dependency between measurements. Thus, we decided to use boxplots to visualize measurements. We also included a basic visualization of engines as a process flow diagram and a correlation matrix view, which was not yet mapped to the boxplot view.

I3 *Causality graph and measurement filtering:* Next, we added the causality graph. First, we included the entire causality graph in which users could switch between a causality graph and a manufacturing process view. The engineers reported, however, that analyzing the entire causality graph was a very difficult procedure. It is for this reason that we only show subgraphs of the causality graph as outlined in Figure 5.
We also introduced a measurement filter. Despite this addition, it was still difficult to see, which measurement in the causality graph belonged to which station in the manufacturing process. We thus added an appropriate hovering event to the causality graph as demonstrated in Figure 6. We also added a scatterplot for attribute pairs in the correlation matrix, to give a more detailed overview of correlations between the measurements.

I4 *Back-end changes:* We subsequently added a manual upload for engines with an error via local CSV files. Furthermore, we changed the back-end of our system to use a SQL database instead of local JSON files as input data for *ManEx*. Thus, users across BMW may use ManEx via a public URL.

I5 *Focus-set validation:* Next, we enhanced

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

the back-end to perform automated dimensionality reduction based on a sample of measurements. As a result of this users were able to build and evaluate *focus-sets*.

I6 *Hypotheses tests and manual engine group selection:* Finally, we introduced hypotheses test and added a manual selection of engines of interest as a new error class. Following on from this, our prime concern was to improve the user experience, for example by mapping selections from the correlation matrices to the boxplot views, or introducing a consistent color scheme for the entire application.

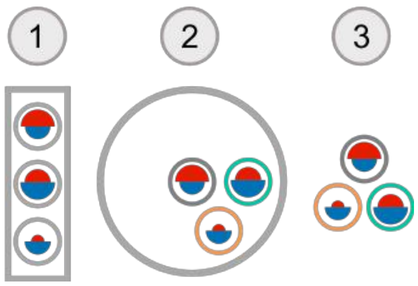


Figure 8: Three different approaches to visualize engine groups. In (1), we show anomalies in a vertical layout, which we adapted in (2) to a radial layout. In (3), we choose to display the glyph for the control group on top and new groups below.

Anomaly Score Evolution

Figure 8 presents three different attempts we used to visualize anomalies for engine groups. In (1), we visualized anomalies vertically. The engineers found, however, that it was a difficult task to differentiate between engine groups, while much space was lost along the vertical axis of our visualization. In (2), we experimented with a radial layout and added a color scheme for engine groups. Here the engineers noted that this visualization looked like a face and the issue of lost space had still not been tackled. Hence, in (3) we chose a more space-efficient layout, where the control group is displayed on top and new groups below.

Causality Graph Evolution

Figure 9 presents four views of the causality graph. In a first attempt in (1), we showed the entire causality graph. Due to a large number of nodes, however, the engineers were confused about where to start their analysis. We thus included an option to filter for the number of nodes in (2). Even with this option, however, the engineers could only filter for sub-graphs with a large number of nodes and leaving small sub-graphs out of the analysis process. Since the engineers were more interested in sub-graphs than in the entire graph, we experimented with different uses of colors and circle sizes in (3) to put the focus on specific regions of interest. This view still failed to show the entire causality graph and the engineers found it difficult to determine where to start an analysis. Thus, in (4) we decided to only show causalities for selected measurements.

Evaluation

Our evaluation aims at validating the usefulness and usability of the proposed visualization design. We present the results of an expert study, embodying qualitative coding of user feedback in combination with a quantitative usability scale.

Participants: We carried out this study together with five automotive engineers from BMW. They were all male between 25 and 35 years old. The small sample size is the result of there only being a few experts with such detailed knowledge about the manufacturing of electrical engines at BMW. This, however, is a common situation when building solutions to very specific domain problems, as has also been demonstrated in other design studies [1], [4], [5].

Data: We uploaded the data of 2985 randomly selected engines, 106 of which had an explainable error and from a period of four months into *ManEx*. The error was known by all participants and they were told that this error was subject to the evaluation. The data for each engine contained 8764 measurements recorded at 154 stations from 54 steps.

Task: We assigned the following task to each participant: *Please find a set of measurements that you believe to be relevant for developing a testing procedure for the specific error*. Since the error was an explainable one, the goal of the task was to identify its root cause early on in the manufacturing process.

Procedure: We performed a think-aloud study [17] with one observer taking notes. First, we gave a detailed system walk-through, where participants were able to ask questions. Next, we

Figure 9: Different attempts to visualize the causality graph. In (1), we showed the entire graph and included a filter in (2). In (3), we experimented with colors and node sizes to put the focus on specific regions of interest in the graph. (4) shows the final visualization, where we only show subgraphs, which are relevant for selected measurements.

asked every participant to perform the predefined task followed by open-ended questions on how each view of *ManEx* supported the execution of the task. Each individual session took 90 minutes on average. To quantitatively assess the usability, we further applied the *System Usability Scale (SUS)* [18] after each think-aloud session.

Results from the Think-Aloud Study

First, all engineers loaded the engines with the explainable errors into *ManEx* (**T1**). Next, they navigated through the hierarchical and temporal dimensions of the manufacturing process using the anomaly glyph representation to compare all engines to the erroneous engines (**T5**). They then added measurements to the *focus-set*, which deviated the most from all engines and had a measurement name, which could be related to electrical errors. Here, one engineer noted *"based on the measurement name I can tell if it makes sense to add it or not"*. Three engineers also used the p-value table to further add measurements to the *focus-set* (**T2**). The other two engineers used the causalities to evaluate similar measurements to the ones they selected (**T6**). The engineers next reviewed the PCA visualization and iteratively added and removed measurements until they were satisfied with the PCA scatterplots. The degree of satisfaction mostly depended on how erroneous engines deviated visually from error-free engines. Three participants used the brush selection to add another outlining engine group (**T4**), where one engineer reported *"I want to see if we have to consider the engines [I selected] as electrical errors"*. Engineers then inspected their *focus-sets* in detail using the boxplots and correlation views. Here, experts again removed measurements from the *focus-set* noting that *"the possibility of evaluating the focus-set in detail is very relevant to validate my assumption on important measurements."* Finally, three engineers encountered one measurement pair in the correlation views, which already helped them to derive a can you describe sophisticated a bit in derail? did they formulate rules or dependencies to match? sophisticated test to detect the electrical error early on in the manufacturing process, where one engineer acknowledged that it is *"very exciting to see how two measurements are already capable of dividing the good engines from the bad ones."* The other two engineers used the high dimensional *focus-set* as a new error identification mechanism.

One engineer suggested including a more detailed description of measurements into *ManEx* but stated at the same time *"It is a little bit unfair to request such features from ManEx because actually, we have to provide this information ourselves when we define the measurements at our stations"*. Apart from that, the engineers had no further suggestions for system improvement.

Results from the System Usability Scale

Apart from the think-aloud study, we found that the usability of our systems with 83.5 is significantly above the average score of 68 [18]. The individual scores are outlined in Table 1.

All the engineers noted that the electrical error could be detected earlier with their interactively created *focus-sets*. This already resulted in a reduction in scraped parts of over 3% at BMW. In

This article has been accepted for publication in IEEE Computer Graphics and Applications. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MCG.2022.3155306

terms of serial manufacturing processes, in which thousands of engines are produced every day, the cost savings would be millions when extrapolated on a yearly basis. Hence, we are confident that *ManEx* will result in similar scenarios, where the causes of errors can be detected early in the manufacturing process.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total
Expert 1	10	7.5	7.5	10	7.5	7.5	10	5	7.5	10	82.5
Expert 2	7.5	7.5	7.5	10	7.5	10	5	7.5	7.5	10	80.0
Expert 3	7.5	7.5	7.5	10	10	10	7.5	7.5	7.5	10	85.0
Expert 4	10	10	10	10	10	7.5	10	7.5	7.5	5	87.5
Expert 5	7.5	7.5	10	7.5	7.5	10	7.5	7.5	10	7.5	82.5
Avg.	8.5	8.0	8.5	9.5	8.5	9.0	8.0	7.0	8.0	8.5	83.5

Table 1: Results of the System Usability Scale [18] with five engineers.

Discussion

We contribute the problem characterization and abstraction for the analysis of a large number of measurements from engines along the *hierarchical* and *temporal* dimensions of a manufacturing process. We can translate some of our study results into more general suggestions, which may guide researcher and practitioners from similar domains.

1) Compute anomaly scores and represent them as glyphs to overcome the hierarchical challenge: The drill-down approach starting from parts down to stations helps engineers in getting a good overview of the manufacturing process. The glyph representation of anomaly scores, for parts, steps, and stations is an easy to understand visualization that facilitates navigating to measurements of interest. Before that, engineers had to manually query and connect data from different data silos and compute anomalies in a slow error-prone process themselves. By contrast, our anomaly visualization allows a fast comparison of engine groups and the evaluation of whether an error can be related to a single part, step, and station or their composition.

2) Identify relations between measurements with t-tests and causal discovery to overcome the temporal challenge: Apart from anomaly scores, t-tests and causal discovery facilitate the identification of relations between different measurements along the *temporal* dimension. In the past, analyses of this kind had to be performed manually by engineers. They now have a method at their disposal for using well established yet powerful measures, which are integrated into multiple linked visualization interfaces for identifying the root causes of errors early in the manufacturing process.

3) Compute residual values and facilitate the creation of focus-sets to overcome the measurement challenge: Formerly, engineers had no means for combining heterogeneous measurements from distinct data sources. The computation of residuals now makes measurements comparable across the manufacturing process. Furthermore, by iteratively adding and deleting measurements to and from the *focus-set*, engineers can immediately evaluate the quality of measurements with the representation of the PCA and the boxplot and correlation views.

Our system design also had some limitations, which we briefly summarize here. First, we only analyzed numerical values and did not include more complex data types, such as time series data. Using,this kind of data poses the challenge that they cannot be visualized as boxplots as easily as single numerical values. A solution to this problem can be to identify the most important sub-sequences in time series and extract single values out of them to visualize them as boxplots.

Furthermore, not all measurements recorded during the manufacturing of engines were available in our study. This is because BMW is currently investing substantial resources in restructuring its data strategy, meaning that not all measurements were available during the time of the study (e.g., missing data of inverters).

Conclusion

We presented a design study on the development of a visualization approach to analyze engines and measurements from a manufacturing process. The resulting VA system *ManEx* provides a clear picture of the *hierarchical* and *temporal* dimensions of the manufacturing process. The glyph representation of anomaly scores allows to *lookup* known engines and measurements. The same anomaly scores allow users to easily *browse* through the manufacturing process and select new interesting engine groups. Boxplot views supported by t-tests and causal discovery help to *locate* measurements of interest. Filtering options and correlation views support the *exploration* of engines and measurements. The representation of anomalies and a consistent color scheme for different engine groups allows the *comparison of engine groups*. *Relations* between measurements can be evaluated in general by using the visualization of the PCA computation and in detail by inspecting individual boxplots of the *focus-set*.

ACKNOWLEDGMENT

This work has been supported by the FFG-COMET-K1 Center Pro²Future (Products and Production Systems of the Future), Contract No.881844. It was also partially funded by the Digital Society Initiative (DSI) at the University of Zurich.

REFERENCES

1. J. Eirich, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. *Transactions on Visualization and Computer Graphics*, 07 2021.
2. M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics*, 18:2431–2440, 2012.
3. F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22:419–435, 2019.
4. M. Sedlmair, A. Frank, T. Munzner, and A. Butz. Relex: Visualization for actively changing overlay network specifications. *Visualization and Computer Graphics*, 18:2729–2738, 2012.
5. L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. Paved: Pareto front visualization for engineering design. *Computer Graphics Forum*, 39:405–416, 06 2020.
6. P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016.
7. M. Wörner, M. Metzger, and T.Ertl. Dataflow-based Visual Analysis for Fault Diagnosis and Predictive Maintenance in Manufacturing. In M. Pohl and H. Schumann, editors, *EuroVis Workshop on Visual Analytics*, 2013.
8. A. Maier, T. Tack, and O. Niggemann. Visual anomaly detection in production plants. In *International Conference on Informatics in Control, Automation and Robotics*, volume 1, 07 2012.
9. J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *Visualization in Data Science at IEEE VIS*, 2020.
10. J. Suschnigg, B. Mutlu, A. Fuchs, V. Sabol, S. Thalmann, and T. Schreck. Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Big Data Visual Exploration and Analytics*, 2020.
11. E. Yen, A. Parameswaran, and W. Fu. An exploratory user study of visual causality analysis. *Computer Graphics Forum*, 38(3):173–184, 2019.
12. A. Detzner, R. Rückschloß, and M. Eigner. Root-cause analysis with interactive decision trees. In *International Conference of Information Visualisation*, pages 322–327, 2020.
13. P. Spirtes, C. N Glymour, R. Scheines, and D. Heckerman. *Causation, prediction, and search*. MIT press, 2000.
14. M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, December 2013.
15. Y. Bishop, S. Fienberg, and P. Holland. *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge, Mass, 1975.
16. Mark Harrower and Cynthia A. Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, June 2003.
17. M. Someren, Y. Barnard, and J Sandberg. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses*. Academic Press, 1994.
18. J. Sauro. Measuring usability with the system usability scale (sus), Accessed 05.07.2021.

Joscha Eirich holds a Master in Information Systems from the University of Bamberg. He is working on his PhD at BMW Group. His research activities focus on the formalization of domain knowledge for machine learning methods.

Georgios Koutroulis, holds a Master in Mechanical Engineering from the Aristotle University of Thessaloniki. He is a researcher at Pro2Future GmbH, located in Graz and pursues his PhD in computer science at the Graz University of Technology. His research focuses on causal discovery in temporal data.

Dominik Jäckle is an independent researcher located in Munich, Germany. He received his Ph.D.

degree in computer science from the University of Konstanz in 2018. His main research interest centers around high-dimensional data analysis.

Belgin Mutlu, is manager of Pro2Future. She holds a master's degree in Telematics and a Ph.D. in Computer Science from the Graz University of Technology. Her research interests include recommender systems and visualization support for big data.

Roman Kern, is the head of Knowledge Discovery, a joint research group at the Institute for Interactive Systems and Data Science at Graz University of Technology and the Know-Center. He was awarded his Ph.D. by Graz University of Technology. Roman's research interest include Machine Learning, Natural Language Processing and Causal Data Science.

Tobias Schreck is a Professor and head of the Institute of Computer Graphics and Knowledge Visualization at Graz University of Technology, Austria. He was an Assistant Professor with the University of Konstanz, Germany, and a Post-Doctoral Fellow with Technical University of Darmstadt, Germany. He received his Ph.D. in computer science from the University of Konstanz in 2006. His research interests include visual analytics and applied 3D object retrieval.

Jürgen Bernard, is an Assistant Professor at the University of Zurich and head of the Interactive Visual Data Analysis group. He has received his Ph.D. in computer science from Technical University of Darmstadt, Germany, and was a postdoctoral research fellow at the University of British Columbia, Canada. His research focus is on visual analytics and human-centered machine learning.

6.5 P5: ManKnowVis: How to Support Different User Groups in Contextualizing and Leveraging Knowledge Repositories

Full reference: *J. Eirich, D. Jäckle, M. Sedlmair, C. Wehner, U. Schmid, J. Bernard, T. Schreck, and J. Bernard. Manknowvis: How to support different user groups in contextualizing and leveraging knowledge repositories. Preprint*

ManKnowVis: How to Support Different User Groups in Contextualizing and Leveraging Knowledge Repositories

Joscha Eirich, Dominik Jäckle, Michael Sedlmair, Christoph Wehner, Ute Schmid, Jürgen Bernard, and Tobias Schreck

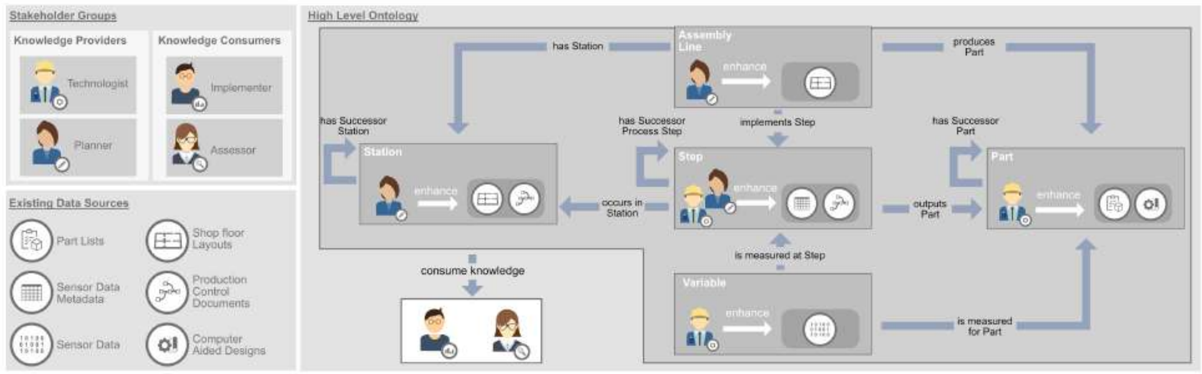


Fig. 1. Relevant stakeholders, data sources, and high-level ontology for ManKnowVis. Technologists and planners provide and implementers and assessors consume knowledge. We use six different data sources as input data for ManKnowVis. A manufacturing process can be described with the entities "Assembly Line", "Station", "Step", "Part", and "Variable". Each entity can be represented with existing data and enhanced by providers. Consumers can inspect this high-quality process knowledge to improve analysis outcomes.

Abstract— We present ManKnowVis, the result of a design study, in which we contextualize data from multiple knowledge repositories of a manufacturing process for battery modules used in electric vehicles. In analyses of manufacturing data, we observed a discrepancy between two stakeholder groups involved in serial manufacturing processes: Knowledge providers (e.g., engineers) have domain knowledge about the manufacturing process but have difficulties in implementing data-driven use cases. Knowledge consumers (e.g., data scientists) have no first-hand domain knowledge but are highly skilled in performing data-driven analyses. ManKnowVis bridges the gap between providers and consumers and enables the creation and completion of manufacturing knowledge. We contribute a multi-stakeholder design study, where we developed ManKnowVis in three main iterations with consumers and providers from an automotive company. The iterative development led us to a multiple linked view tool, in which, on the one hand, providers can describe and connect individual entities (e.g., stations or produced parts) of the manufacturing process based on their domain knowledge. On the other hand, consumers can leverage this enhanced data to better understand complex domain problems, thus, performing data analyses more efficiently. As such, our approach directly impacts the success of data-driven use cases from manufacturing data. To demonstrate the usefulness of our approach, we carried out a case study with seven domain experts, which demonstrates how providers can externalize their knowledge and consumers can implement data-driven use cases more efficiently.

Keywords: Design study, Knowledge-Assisted Visualization

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces (GUI); User-centered design

1 Introduction

Visualization has an intrinsic motivation and long-standing tradition of applied and problem-driven research [25]. The typical goal of such projects is to enable domain experts to understand and analyze their own domain data through carefully designed visualization interfaces [41]. In other words, data expertise is brought to domain experts through interactive visualization. While this scenario still resonates well with many classic scientific setups, with the advent of data science now also collaborative setups are becoming more common, in which data scientists and domain experts collaborate toward a joint solution. Instead of enabling domain experts directly, the main challenge in such setups lies in bridging the knowledge gap between differently-skilled stakeholder groups. We argue that visualization plays a key role in bridging the gap between stakeholders with different levels of knowledge and we report on an example of such a multi-stakeholder design study from the automotive industry.

This industry currently faces fundamental changes due to the digitization of its production processes [20], where data-driven analyses are already firmly integrated into industrial manufacturing processes [19]. Recent success stories show how such analysis efforts result in substantial manufacturing improvements. Examples are the visual exploration of assembling line performance to detect inefficiencies [54], the analysis of acoustic signatures of electrical engines to improve manufacturing quality [9], or the identification of patterns in machine repair logs to decrease maintenance costs [14]. When working in this area, we noticed a discrepancy between two stakeholder groups, those that *pro-*

- *Joscha Eirich, Christoph Wehner, and Ute Schmid are with University of Bamberg. Email: firstname.lastname@uni-bamberg.de*
- *Tobias Schreck is with Graz University of Technology. Email: tobias.schreck@cgv.tugraz.at*
- *Jürgen Bernard is with University of Zurich. Email: bernard@ifi.uzh.ch*
- *Dominik Jäckle and Joscha Eirich are with BMW Group. Email: firstname.lastname@bmw.de*
- *Michael Sedlmair is with University of Stuttgart. Email: Michael.Sedlmair@visus.uni-stuttgart.de*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

vide knowledge, and those that *consume* it. Knowledge *providers*, such as engineers, have domain knowledge about specific parts of the manufacturing process. For instance, they know exactly what data is recorded at manufacturing stations. However, *providers* do not have the necessary skills to perform sophisticated data-driven analyses (hereafter: **use cases**), such as the training of machine learning models. Such use cases are usually performed by highly skilled data experts, such as data scientists. However, this group does not have first-hand domain knowledge about the manufacturing process. For example, we interviewed seven data experts, who reported that for previous use cases they required knowledge about a robotic arm location inside a manufacturing station that might contain an error or about the interconnection of battery cells for different derivatives of battery modules. Therefore, they rely on consuming knowledge, which is why we refer to them as *consumers*.

Instead of enabling *providers* directly via visualization interfaces, the main challenge in manufacturing setups is to bridge the knowledge gap between these two groups and foster fruitful, effective, and seamless collaborations. While such setups have been rarely reported in visualization studies so far, we argue that – with the advent of the era of data science – these setups will become increasingly common and thus also important for the visualization community. In fact, we argue that visualization can play a key role in such collaborative setups, with the main intention to foster knowledge transfer between differently-skilled stakeholder groups. The goal of our work is to report an example of such a multi-stakeholder design study in collaboration with BMW to illustrate the value that visualization can add to such scenarios. In this regard, we report the challenges *consumers* face when accessing manufacturing knowledge and how *providers* can leverage visualization interfaces to externalize their knowledge.

To access knowledge about use cases, *consumers* have two options: First, they can ask *providers* to provide them with information about a use case. Second, *consumers* can consult documentation about the manufacturing process. *Consumers*, however, face the following challenges when accessing knowledge from *providers* or existing documentation:

In theory, *providers* can support *consumers* with relevant knowledge about use cases. However, it is often not clear, how to find *providers* to answer specific questions about use cases. Furthermore, *providers* and *consumers* reported requiring a long time to find a common language as the foundation for a discussion. *Consumers* can also study existing documents about the manufacturing process, such as process flow diagrams. However, these documents are often not self-explanatory. Thus, even though rich valuable repositories of knowledge exist, *consumers* find it hard to access and comprehend these repositories. As a response, we propose ManKnowVis (Manufacturing Knowledge Visualization), a *knowledge-assisted visualization* system [12] that resulted from a design study [41] project in collaboration with *providers* and *consumers* at BMW.

We base our study on existing documentation from different data sources, which provide a holistic depiction of an assembly line for battery modules from one of BMW's manufacturing facilities. Furthermore, we were provided with an existing ontology that supported us in contextualizing entities of the analyzed assembly line. *Providers* can use ManKnowVis to enrich existing manufacturing entities with their knowledge by describing and connecting entities. ManKnowVis provides them with visual interfaces to externalize their knowledge [33,51], such as grouping stations via annotation functionalities directly on shop floor layout images. *Consumers* can use ManKnowVis to get a detailed understanding of distinct use cases they are addressing. For example, *consumers* can inspect annotations to answer questions such as *how are different stations connected* or *what property of a product do specific variables measure?* Therefore, ManKnowVis bridges the gap between *providers* and *consumers* by supporting *providers* in externalizing their knowledge in a systematic and structured way and by providing *consumers* with efficient access to contextualized knowledge, which was previously inaccessible. For *consumers* this knowledge is particularly helpful in understanding complex manufacturing data, which is the necessary foundation to perform use cases, such as root cause error analyses or anomaly detection to name a few. As a result, ManKnowVis directly influences the success of use cases.

In summary, we contribute (1) the problem characterization and abstraction about contextualizing and accessing knowledge for a manufacturing process; (2) the design of ManKnowVis, which supports *providers* in externalizing their knowledge and *consumers* in accessing previously inaccessible knowledge; and (3) the evaluation of ManKnowVis with a case study and reflections of our design process.

2 Related Work

We start by giving a brief summary of related work about knowledge assisted visualization (Section 2.1) and its applications (Section 2.2), followed by design studies in the automotive sector (Section 2.3).

2.1 Knowledge-Assisted Visualization

The incorporation of knowledge through knowledge-assisted visualizations into analytical tasks has been of great importance to information visualization [5, 24, 29, 51]. Knowledge can be divided into two types: *explicit* knowledge that can be externalized easily, for example, in words or numbers; and *tacit* knowledge that is inherent to the individual [11, 12, 33, 51]. The latter is often not recognized by the individual as knowledge but rather expressed through action, commitment, and involvement, which renders it notoriously difficult to externalize [12].

With the capabilities of generating and externalizing knowledge from knowledge-assisted visualizations, domain experts can perform more efficient and effective data manipulation and analysis tasks [49–51]. Tacit knowledge from such analyses can be externalized and stored, for example via rules [30] or labels [10]. The resulting explicit knowledge can then be transferred among different users providing a common context that helps to bridge differing backgrounds and to reduce the burden to acquire domain knowledge about different domain problems [29].

Furthermore, different knowledge types provide different support when leveraged in knowledge-assisted visualizations [24, 29, 49]. Here, *operational knowledge* deals with knowing how to use visualization systems, while *domain knowledge* contains knowledge about analyzed data helping users to correctly interpret visualizations. In our design study, we are particularly interested in how *domain knowledge* can be formalized and combined with new *explicit* knowledge to support users with less domain knowledge.

2.2 Applications of Knowledge-Assisted Visualizations

Several researchers have made noticeable efforts to capture *explicit* and externalize *tacit* knowledge via different visualization techniques. Previous knowledge-assisted visualizations were based on *rules* [30, 53], *metadata attributes* [32], *graphs* [15, 27, 44, 45, 56], *labeling* [3, 17, 36, 52], or *annotations* [35] of data instances. We review examples of these different approaches in the following.

Regarding *rules*, Moritz et al. [30] implemented a VA system that externalizes design knowledge as a collection of rule-based logic sets to accelerate system designs, while Xiao et al. [53] externalize knowledge via rules for the visualization and classification of network traffic. In the case of *metadata attributes*, Nie et al. [32] support physicians in the classification of breast lesions by providing additional metadata attributes for each lesion. Kadlec et al. [18] build a system that incorporates *tacit* knowledge of geologic experts to support the segmentation of seismic data sets through metadata attributes. An example for *graphs* is provided by Stitz et al. [45], who externalize *tacit* knowledge in the form of visualization states and analyses with a provenance graph. Sobral et al. [44] display *ontologies* as graph to formalize knowledge from urban mobility events. Regarding *labeling*, Berg et al. [3] provide a system for capturing expert knowledge via labeling bio images. In the case of *annotations*, Ren et al. [35] built a system that allows users to augment charts via different annotation interactions.

All presented studies successfully demonstrate how to capture *explicit* and/or externalize *tacit* knowledge. In our design study, we aim to leverage both knowledge types to support *consumers* with use cases. While previous approaches relied on one way to externalize knowledge, we combine multiple techniques implementing different ways to formalize knowledge. Particularly inspiring for the specific design of ManKnowVis were the formalization of expert knowledge via *ontologies* [44] of manufacturing processes, and the enrichment of *explicit* knowledge via externalizing *tacit* knowledge from different groups of domain experts via *metadata attributes* [18,32] and *annotations* [35] for entities, such as steps, stations, or parts from a manufacturing process.

2.3 Design Studies in Automotive Industry

Visualization to date has contributed substantially to help to analyze complex data in manufacturing settings, as the recent survey by Zhou et al. [57] shows. A key domain is the automotive sector. There, design studies and resulting VA applications mainly support product design, condition monitoring of stations, the optimization of testing procedures, or the visual support of high cognition tasks. Efforts were carried out to visualize in-car communication networks [38–40], to facilitate the exploration of multi-criteria alternatives for rotor designs [6], to detect and analyze anomalies in test stations [10,48], the visual exploration of assembling data to detect inefficiencies [54], and to support mechanical engineers in the analysis of acoustic signatures of electrical engines [9].

Some of the mentioned studies explicitly acknowledge the need for externalizing *tacit* expert knowledge [10]. For example, the *Cardiogram* system [40] stores externalized expert knowledge in the form of state machine diagrams, while *IRVINE* [9] stores expert knowledge in the form of labels for electric engines and annotations in the raw sensor data. All mentioned studies succeeded in creating insights for engineering experts based on machine sensor data. The problem we face, however, is how *consumers* with a much lower level of engineering knowledge can leverage expert knowledge. Grounded on previous findings, we built a system that processes, combines, and contextualizes different *explicit* knowledge repositories, which are enriched with *tacit* knowledge from multiple *providers*. The resulting combined knowledge is available for *consumers* through visual interfaces, which supports them in better comprehending use cases.

3 Methods

During this study, we primarily followed Sedlmair et al.'s nine-stage framework for design studies [41]. In addition, we used the *Nested Model* for visualization design and validation by Munzner [31], which guides a more detailed problem characterization, the data operation and abstraction, the visual encoding and interaction design, and the algorithm design.

To understand the domain problem, we performed online interviews with seven *consumers* and seven *providers*. We asked *consumers* questions about how they perform use cases, what kind of challenges they face, or which information they need during such projects. Regarding the *providers*, questions were about what information is generally important when analyzing manufacturing data or what are general documentation tasks they perform in their daily business. The interviews helped us in understanding, contextualizing, and abstracting the domain and the domain problem of missing knowledge during use cases.

To describe production processes and involved stakeholders, we relied on an existing ontology from BMW. The ontology guided us in developing ManKnowVis and representing its underlying manufacturing data. A detailed description of the ontology and how *explicit* and *tacit* knowledge sources are embedded into it, is provided in Section 5.1.

Our system development went through three main iterations over five months, during which we interviewed *providers* and *consumers*, tested design alternatives, and held discussions with visualization experts. Each iteration was carried out in close collaboration with two *providers* and *consumers*. They accompanied the system development with knowledge about the data characterization (Section 4.3), and resulting tasks (Section 5.3). They also gave a constant stream of feedback on the visual design of our system. The exchange with each user took place in up to two meetings per week ranging from 30-90 minutes. During these meetings, fundamental characterization and design aspects were discussed and open issues were clarified. We evaluate ManKnowVis with a case study. Methodological details for this downstream evaluation are provided in Section 8.

4 Problem and Data Characterization

We report the characterization of the problem domain about transferring knowledge between *providers* and *consumers*. First, we provide an introduction to use cases in the automotive manufacturing sector (Section 4.1), followed by a description of the problem our collaborators face (Section 4.2) and the data we included (Section 4.3).

4.1 High-Level Description of Use Cases

To better comprehend the domain of contextualizing knowledge of manufacturing data, we analyzed several data-driven use cases at BMW. Examples include the identification of error causes by connecting and analyzing sensor data across different steps, the development of visualization approaches for condition monitoring at individual stations, or the optimization of cyclic times to increase manufacturing outputs. We interviewed seven *consumers* with different roles, such as coordinating use cases, performing data analyses, or building data infrastructures. In addition, we interviewed seven *providers* responsible for the planning of manufacturing processes or specific stations in an assembly line. After interviewing the *consumers*, we observed that they could be divided into two main groups that are involved in carrying out use cases; one responsible for the assessment of use cases and one for implementing them. They can be described as follows:

Implementers are required to carry out use cases. They are highly skilled at data related tasks, such as data visualization (e.g., the development of dashboards for manufacturing data), data engineering (e.g., setting up data infrastructures to store machine sensor data), or sophisticated data analyses (e.g., training machine learning models to predict manufacturing output quality). They are interested in specific information about the manufacturing process, for example, the functionality of a single station or where to find variables for a step.

Assessors evaluate the benefit of a use case, prioritize different use cases, and assign *implementers* to use cases. For them it is important to have a high-level overview of the manufacturing process, for example, to ensure that data from all manufacturing stations are recorded properly.

Even though *consumers* are skilled in the analysis of manufacturing data, they need to acquire additional knowledge to realize use cases. Examples of such knowledge include information about produced products (e.g., Is the analysis object a battery cell or an electric engine?) or about the manufacturing process (e.g., Does an error occur at a testing station or an assembly station?). To acquire a sufficient degree of knowledge, *consumers* have two options:

(1) *Investigate existing data sources*: Much information about the manufacturing process is already documented. To gather further information, *consumers* manually analyze existing data sources that are stored in the form of documents on local file drives, or across organizational databases.

(2) *Request support of providers*: *Consumers* directly approach *providers*, who have the necessary knowledge about parts of the manufacturing process. In our design study, we surveyed that for use cases the most relevant *providers* are planners and technologists, which can be described as follows:

Planers have general knowledge about the manufacturing process as a whole integrated system. They know where stations and steps are located, how they are connected, and how parts are flowing through assembly lines.

Technologists have detailed knowledge about individual steps in the manufacturing process. They know what individual stations do, what variables they measure, and what kind of parts are produced.

4.2 Domain Problem

The challenge of accessing manufacturing knowledge is present across the manufacturing sector [6, 9, 28, 39]. As well as many companies, BMW is also investing large resources in the digitization of its processes. In this regard, *consumers* face the following challenges when investigating existing data sources or interacting with *providers*:

Investigate existing data sources: The interviewed *consumers* reported that due to the complexity of use cases, the documentation in most of the cases is not self-explanatory. As one implementer noted *"I had a case where several recordings were related to the same variable. Only after consulting a colleague, I found out that this was because the values were actually points on a time series curve of a variable."*

Another challenge from existing data sources is that *consumers* often do not know where to find relevant information. Furthermore, information is fragmented across multiple tools, which are not integrated. In this regard one assessor mentioned that *"sometimes documentation can be scattered across different databases. Hence, we sometimes have difficulties finding out if we have all relevant information at our disposal."* Furthermore, *consumers* and *planners* reported that documentation is often not complete as one implementer mentioned *"sometimes we have difficulties in finding out if the data we are analyzing is up to date."* Finally, *providers* reported that in they *"do not have a system, which uses data from more than one database. Having such a system at hand would ease documentation efforts for the manufacturing process"*. Thus, they often use individual documentation solutions, such as local files, which makes it even harder for *consumers* to access this documentation.

Request support of providers: *Consumers* and *providers* reported that due to their background they also have different mental models. Here one technologist explained that *"especially when use cases start we do not have a common language. I remember that we had a kick-off meeting, where all participants needed more than 40 minutes until it was clear what we were actually talking about"*. Furthermore, *providers* are often hard to reach, where one assessor noted *"we had a project where timelines between us and engineers were challenging to align risking a delay"*.

Due to these challenges is especially relevant *consumer* to efficiently access knowledge about the manufacturing process. In turn, *providers* need a way to easily support their documentation efforts.

4.3 Used Data Sources

Providers and *consumers* reported that they consider the following data as important to give a holistic overview of the manufacturing process.

Production Control Documents: In these documents, planners describe manufacturing steps (e.g., welding or gluing) and the sequence of how stations perform these steps. This information is relevant to knowing what is done at stations and how they are related to each other.

Sensor Data Metadata: When conceptualizing an assembly line, technologists specify what variables are recorded at stations. This includes information about variables (e.g., what unit the variable has or at which step a variable is recorded). That information is relevant to knowing, what kind of variable is recorded and how it can be interpreted.

Computer aided designs (CAD): When designing new products, technologists produce various CADs, which are images showing the products. This information is relevant to understanding how the parts - being the core object of many analyses - look like.

Shop floor layouts: Each manufacturing plant contains various shop floors, where parts are produced and assembled. Here, planners arrange stations in shop floor layouts. This information is relevant to review at which stations parts are assembled into new parts and how stations are related.

Part lists: When designing new products, technologists do not only provide CADs for individual parts, such as battery cells but also information on how parts are arranged to form another part (e.g., a battery module contains battery cells and heat-conducting plates). This information is relevant to understanding how parts are related to each other.

Sensor data: Technologists specify what kind of data must be recorded inside each station, to monitor either the station itself or the parts it produces (e.g., multiple temperature or voltage recordings).

5 Abstractions

Based on the domain problem and the used data sources we provide further abstractions. First, we relate users and data to an existing ontology (Section 5.1). Building upon the ontology, we derive requirements for ManKnowVis (Section 5.2). The requirements helped us in the abstractions of our tasks, which aim to support *providers* in externalizing their knowledge and *consumers* in accessing this knowledge (Section 5.3).

5.1 Data Characterization

The high-level ontology, depicted in Figure 1, outlines how we abstracted from existing data sources and how users interact with such data sources. The ontology comprises five entities (assembly line, station, step, part, variable) and serves to describe an assembly line of a manufacturing process. An assembly line has stations, implements steps, and produces parts. Steps, such as gluing or welding, occur in stations and output parts, such as battery modules or battery cells. To represent the sequence of steps, stations, and parts, these entities can have a successor. For example station 2 can follow station 1 or a battery cell and an insulating film form an insulated battery. Variables are measured in steps and help to evaluate parts. For instance, resistances show if the welding of a battery module was executed correctly.

The existing data sources (see Section 4.3) already provide a description of entities and their relation in the ontology. For example, part lists indicate how parts are assembled into new parts. Since the existing data may be incomplete, *providers* can enhance each entity in light of their own knowledge. For example, a planner can use ManKnowVis as a visual editor to arrange stations on a shop floor layout or technologists can assign a unit to a variable. While *providers* enhance each entity of the ontology, *consumers* can inspect each entity and the relations between entities to better understand aspects of the manufacturing process. For example, an *implementer* can analyze what variables belong to a step or an *assessor* can check in which step a part is produced. This high-level ontology provided us with a framework to derive overarching system requirements and to abstract tasks for each entity separately. Please note that the ontology BMW uses is much more detailed and contains more specific entities. Due to non-disclosure agreements, we are required to use this high-level ontology.

5.2 Requirements

Providers have a higher level of knowledge about the manufacturing process than *consumers*. To better analyze and assess this knowledge gap, we structure the requirements based on levels of knowledge about entities in the ontology (See Figure 1), which we refer to as **knowledge maturity levels**. To compute maturity degrees for each entity, we were guided by the available data sources (See Section 4.3). For example, the shop floor images include the location of stations but no information about part flows. As such, we can use shop floor images to describe stations but not parts. As a result, we derived the necessary information for each entity in Figure 2, which is necessary to achieve a sufficient *knowledge maturity level*.

Fig. 2. Requirements to increase maturity levels of ontology entities. Entities either need a description or must be related to other entities.

Entities can either be provided with a **relation** or a **description**. In terms of entity descriptions, a station needs a location on the shop floor and a sub-line. A sub-line defines a cluster of stations that all can be assigned to an overarching process. For example, stations 1, 2, and 3 could all be responsible for preparing lithium-ion cells. Parts, variables, and steps all need a description, while variables further require a unit. Steps need also a sub-process. A sub-process defines different sub-steps that are executed within a step. For example, the painting of a lithium-ion cell can be divided into the sub-processes painting side A and painting side B. Furthermore, entities need a relation to other entities. For example, a variable can be measured for a part, a station can have a successor, or a step can be executed by multiple parallel running stations.

As a result, we can compute the *knowledge maturity* for each entity by checking what information is necessary to fully enhance it. In Figure 2, for instance, a variable contains two possible description fields and could be related to one part. If a *provider* adds a description and a unit, the *knowledge maturity level* is 66%. This leads us to one overarching requirement (R) for each user group. For the providers, the requirement is to *enhance the knowledge maturity of entities* (RP). In turn, *consumers* require to *increase their knowledge about entities (RC)*.

5.3 Task Abstraction

We present the abstracted tasks that serve as primary design targets for ManKnowVis. *Providers* and *Consumers* interact differently with ManKnowVis, which is why we differentiate between their tasks below:

Providers can enhance data sources (*RP*) with the two tasks:

P-T1 Describe Entity: Providers need to describe each entity in the ontology. They can describe entities in multiple ways, such as text inputs or annotations.

P-T2 Connect Entity: To define the relationships among entities, *providers* also need to connect different entities with each other. Entities can be connected by assigning different kinds of entities to each other, such as a step to a product or by connecting entities with the same type, such as a station to a successor station.

Consumers can increase their knowledge (*RC*) with three tasks:

C-T1 Overview of Entities: Consumers need to locate relevant data instances in the manufacturing process, for example, where a step of interest is executed. Furthermore, *consumers* need to evaluate how entities are related, such as which stations belong to a part.

C-T2 Drill-down and Expand: Consumers exploring large numbers of entities need support for the drill-down to entities of interest. The information need may differ between the location of stations or the distribution of measured variables. After an entity of interest is identified it needs to be evaluated in the context of other entities, for example, what other stations also measure variables of the same unit. Thus, *consumers* also require support to expand back to an overview.

C-T3 Hypothesize: While interacting with ManKnowVis, *consumers* are constantly able to confirm or reject hypotheses. Such hypotheses can regard single entities, for instance, *what unit does a variable of interest have*, or multiple distinct entities, such as *how is the data quality of all variables for a product*.

6 The ManKnowVis System

We present the design and implementation of ManKnowVis, with an overview of its main views in Figure 4. ManKnowVis comprises three views that represent three levels of detail [42] of the manufacturing process. A general overview is shown in the "Parts and Steps" view. A middle level of detail is displayed in the "Stations" and the highest level of detail in the "Variables" view. We use the notation A-C to refer to the views and sub-components (e.g., A1, A2, A3) later on. Two views are always displayed side by side. Users can switch between the three views using the top panel in Figure 4. For the views, we used the showcase data-set from an assembly line for battery modules containing the described data sources from Section 4.3. Due to non-disclosure agreements with BMW, we anonymized all data and cannot mention details, such as how many stations exist or how many parts were produced.

6.1 Overview of the System

(A) shows an overview of parts and steps of an assembly line (*C-T1*). We represent part lists with a tree layout (A1) [22]. Some parts, such as battery cells can contain a CAD, while other parts, such as glue, do not have CADs. Selecting a part with a CAD results in a details dialog, as shown in Figure 3.

(A2) shows the sequence of steps that can be executed by multiple stations in parallel. We display the sequence of steps in a Sugiyama layout [8], which is specifically designed to display node-link diagrams as a hierarchy and avoids edge crossings. We draw the sequence of steps on the horizontal axis from left (start of the assembly line) to right (end of the assembly line). On the vertical axis, we draw the stations. so that different stations can be shown in parallel (*C-T1*). Sub-lines are shown as gray rectangles in (A2). Stroke colors indicate the step type (pink: testing step, blue: assembly step, yellow: reworking step). Knowledge-maturity levels, for example (A3), are shown as donut chart icons (*P-T1/2)* in all views, where the green area indicates the completeness of an entity. Hovering over the donuts displays a tooltip indicating to which degree an entity is completed and which information is missing. As outlined in Section 5.2, we compute the knowledge maturity degree by inspecting how much information for each entity is provided according to Figure 2. As soon as *providers* enhanced an entity (*P-T1/2*), its *knowledge maturity* donut updates in each view. *Providers* and *consumers* can filter for maturity levels using the sliders at the top right of (A1), (A2), and (B). Details for parts and steps are shown in (A4) (*C-T2*). For example, the description of a step is displayed as a text area and sub-processes as a table.

(B) depicts the stations of an assembly line as a middle level of detail (*C-T1)*. In the background, we show the real shop floor image and display stations as a node-link diagram (B1) on top of it. The computation of locations of stations on the image is outlined in detail in Section 6.3. The position of station rectangles is determined by their original position on the shop floor where we use the same color coding as in (A). We define stations that are available on the shop floor but not in other data sources as *missing stations* and display them as gray rectangles. Sub-lines are shown as gray rectangles in (B2).

Selecting a part or step from (A) or a station from (B) depicts its variables in (C) (*C-T2*). Metadata attributes for a variable are displayed as a table in (C1) and its distributions in (C2). We show distributions on a grid and differentiate between the following three variable types: Categorical variables, (e.g., *status ok* vs. *status not ok*) are displayed as donut charts. Numerical variables (e.g., one temperature recording for multiple parts) are displayed as a histogram. Sequence variables (e.g., multiple numerical recordings for multiple parts) are displayed as a sequence of boxplots. Regarding histograms, we use 20 bins according to the empirical studies and guidelines of Shanann et al. [37]. To inspect the same variable for parallel running stations, users can add additional station recordings to each grid cell with the colors green, orange, and blue. In our design study, we never encountered situations where more than three stations recorded the same variable. To overcome visually overlapping variable recordings for parallel stations, users can filter for single stations by selecting individual stations in (C3). Zooming and panning (*C-T2*) ease the navigation of the views (A1, A2, B, and C2).

6.2 Parts and Steps

The view of parts and steps in (A) is particularly designed to provide an overview of the manufacturing process (*C-T1*). By selecting a part in (A1) and a step in (A2), *consumers* are provided with additional entity details (*C-T2*) in (A4), such as sub-processes for a step in a table view. A drill-down to details is further supported by inspecting CADs of a part as shown in Figure 3.

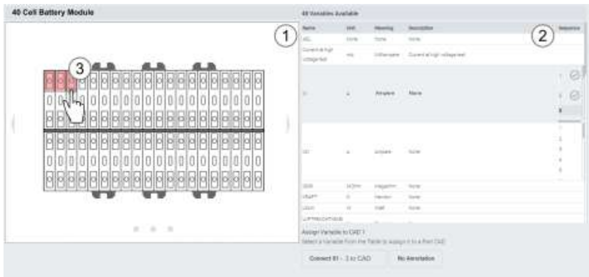


Fig. 3. Details dialog for parts. (1) shows a part's CAD and (2) its variables. Providers can connect (2) with (1) by adding an annotation in the form of an adaptable rectangle as shown in (3).

This dialog shows available CADs for a part with a carousel view on the left hand side. Variables and their attributes are displayed on the right (See Figure 3 (2)), where sequence variables contain an additional column "Sequence". This dialog supports *consumers* in answering questions, such as *how does a part look like* or *what variables are measured for a part* (*C-T3*). In the case of steps, *consumers* are supported in finding out how steps are related or to what sub-lines they belong to (*C-T3*). *Providers* can add a description (*P-T1*) for

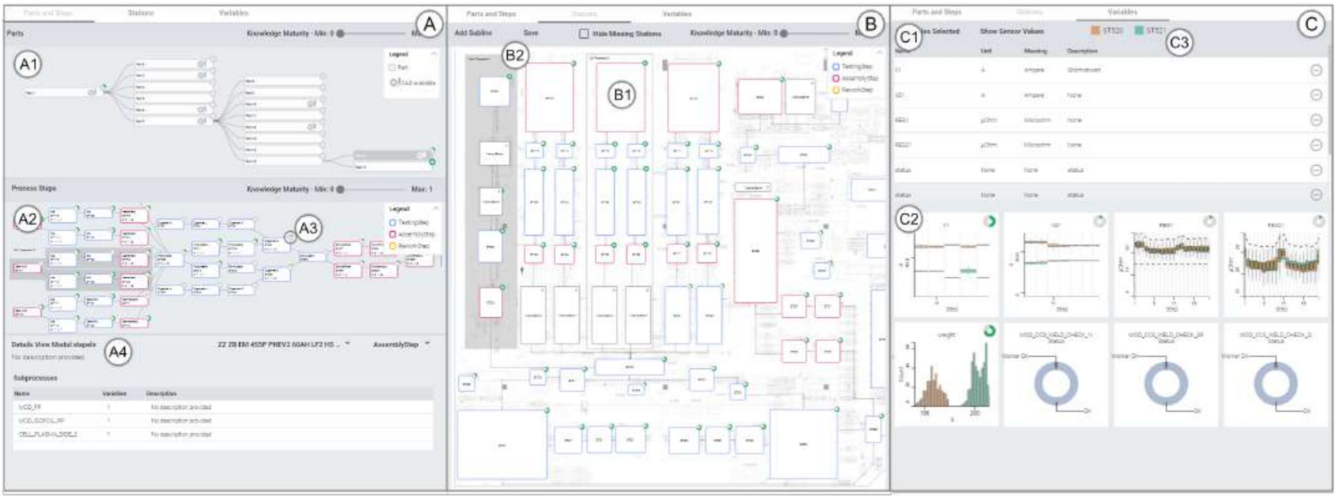


Fig. 4. ManKnowVis is divided into the three level of detail views, high, middle, and low. The *parts and steps* view (A) shows the coarsest, the *station* view (B) the middle, and the *variable* view (C) the highest level of detail. On the user's desktop screen two views are shown side by side, where users can freely switch between each view using the navigation panel on top. (A) shows parts (A1) and steps (A2) and details for both entities in (A4). Knowledge maturity levels are shown as donut charts (A3) in the views A, B, and C. Stations are displayed in (B) and arranged following the physical layout of an assembly line (B1), which is shown as a background image. Sub-Lines are displayed in (B2). To save display space, details for stations are also displayed in (A4). Selected variables for a part, step, or station are displayed in C. Metadata variable attributes are shown in (C1) and their distribution in (C2). Depending on their type, variables are displayed as histograms, multiple boxplots, or donut charts.

parts and steps in the details view with a text field on the top of (A4). Furthermore, *providers* can assign a part to a step (*P-T2*) and change the type of a step, such as an assembly step to a testing step with drop-down selections on top of (A4). Furthermore, *providers* can assign variables to a part (*P-T2*) in the CAD dialog (Figure 3). This is done, by selecting a variable from the table view and dragging a rectangle over the part (See Figure 3 (3)). If a variable cannot be connected to a part via an annotation, *providers* still can assign the property *no annotation* to a variable. Annotations are valuable to *consumers* since they enable them to relate variables not only to a part but also to a specific region of interest on the part (*C-T3*).

6.3 Stations

The station view in (B) represents a middle level of detail and provides information about the assembly line and its stations. The image of the shop floor already contains station names and locations embedded in the image. However, to enable interaction with individual stations and their linking to other views, we also encode them as rectangles on top of the shop floor image. As a first step, we extract the x and y coordinates, as well as the width and height of each station rectangle from the shop floor image using computer vision. To extract the coordinates, as well as the height and with, we apply the library *OpenCV* [7]. Next, we extract the name of each station from written strings on the image with the neural network *tesseract* [43].

With this initial data prepossessing, *consumers* get an overview of stations and their relationships (*C-T1*). Since in an assembly line, stations can be far apart, we highlight successor stations when hovering over a station (*C-T1*). To save space, we display details for selected stations also in (A4). *Consumers* can use this view to answer questions like *Where are stations located* or *how are stations related* (*C-T3*).

Providers can enhance stations in three ways. First, the position of a station can be changed (*P-T1*) by dragging a station rectangle. Width and height can be adapted by dragging its borders. Second, *providers* can relate stations with each other by drawing links between them (*P-T2*). This is done by first clicking on a rectangle and then dragging a line to another rectangle. This results in a straight line between the two stations. However, in the real assembly line, station links are not always straight lines. Hence, we allow to freely adapt each line by setting individual breakpoints on lines. Third, *providers* can click on "add Sub-line", to add another sub-line (*P-T1*), which is as well as stations is a freely adaptable rectangle in terms of x and y coordinates and width and height. By dragging a sub-line over a station, the station is automatically assigned to the sub-line. *Providers* can change the name of a sub-line by clicking on the name at the top left of each sub-line rectangle and inserting a new name in a dialog field.

6.4 Variables

The variable view in (C) is designed to show the highest possible level of detail. Variables can be displayed by selecting a part in (A1), a step in (A2), or a station in (B). All selected variables are displayed in a table view in (C1) (*C-T1*). Besides the name of a variable, the table view also includes metadata attributes, such as the unit, the meaning of the unit, and a description of the variable, which serves to better comprehend the meaning of a variable (*C-T2*). The sensor data values of variables are shown in a grid view in (C2). By visualizing the distributions of the sensor values, *consumers* already get an overview, of which variables may serve well for use cases (*C-T3*). By inspecting distributions of numerical and sequence variables, *consumers* can for example answer questions like *which kind of model could be appropriate for the available data* or *what variables are suitable to extract features for a model?* By analyzing categorical variables *consumers* can answer questions such as *are there enough labels to train a model* (*C-T3*)? In terms of enhancing variables, *providers* can edit attributes in the table view, by entering units, unit meanings, and variable descriptions *P-T1*.

6.5 Linking of Views

In addition to inspecting individual views, we provide a linking between all three levels of detail views.

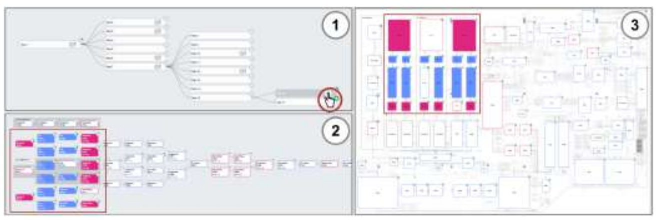


Fig. 5. Linking of one entity to related entities. For example, hovering over a part (1) highlights its steps (2) and stations (3).

Hovering over an entity in each of the views, highlights related entities in the other two views. An example is provided in Figure 5. Here, a battery cell is selected in (1) and according steps are highlighted in (2) and stations in (3) (red outline in Figure 5). Hence, *consumers* can answer questions, such as *what stations are involved in the assembly of one selected part*, or *where in the assembly line is a specific step executed* (*C-T3*). By doing so, we aim to implement a drill-down and expand workflow, where *consumers* select an entity from one view, drill down to an entity of another view and then again expand to a different view to analyze how this entity is related to other entities (*C-T2*). The same linking also supports *providers* in identifying where to find entities that must be further enhanced (*P-T1/2*). For example, Figure 5 shows what stations in (3) belong to a battery cell from (1). However, three stations are not highlighted in (3) and hence need also be assigned to a part.

Furthermore, variables can be linked to parts via annotations, which is demonstrated in Figure 3 (3). If variables are linked to parts, we display this information as a tooltip. By hovering over histograms or columns of boxplot sequences in (C2), the according CAD and the provided annotation is shown as demonstrated in Figure 6. This relation allows *consumers* to immediately see to which exact region of a part a variable belongs (*C-T3*).

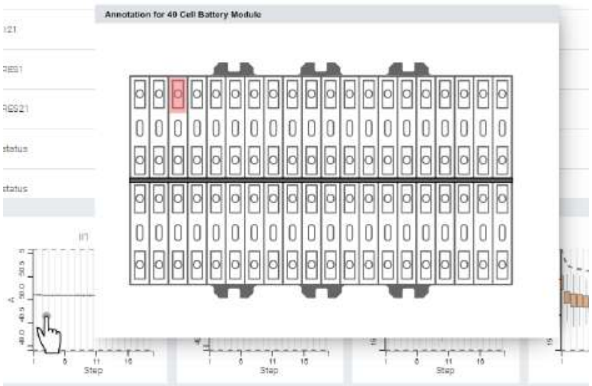


Fig. 6. Tooltip to show the relation of a variable to a part. In the example, hovering over the third column of the variable sequence shows the location of the measurement on the battery module.

7 Design Iterations

ManKnowVis was designed and developed during three main iterations in collaboration with *planners* and *consumers*. During each iteration, we gathered feedback and adapted our visualization design. In the following, we will describe each iteration in detail:

Iteration 1: At the beginning of the project, we aimed to develop a single view capable of visualizing the entire manufacturing process. To do so, we experimented with a force-layout [34,46], which is shown in Figure 8 (1). In this approach, we focused on visualizing steps as a sequence, where green and purple strokes indicated the step types (assembly and testing step). In the same graph, we also represented stations with a blue stroke (1a) and the assembly line with a red stroke (1b). The representation of parts was not yet included. Selecting a step from the graph unfolded its stations (See red outline in Figure 8 (1)). However, *consumers* and *planners* reported that displaying all entities in a graph was confusing since they all have to be interpreted differently. Furthermore, users reported that edge crossings made the results incomprehensible. Regarding our color coding, one of our collaborators reported having color deficiencies and was not able to distinguish between different entities.

Iteration 2: In the second iteration, we first adapted our color scales to account for color deficiencies [26]. Next, we aimed to show the sequence of steps and stations in a more organized manner avoiding edge crossings. Instead of focusing on steps, we intended to display a sequence of nodes containing high-level information about parts, steps, and stations of an assembly line (*C-T1*). On the top of each node, we showed the step name, then stations, and below that the number of sub-processes (S) and variables (V).

Parts (P) were displayed as gray rectangles on the right side of a node. As outlined in Figure 8 (2), we displayed nodes as a sequence with line breaks. To show additional levels of detail (*C-T2*), we provided a tooltip as shown in Figure 7. Here, we displayed the step description in (1), stations as rectangles in (2), subprocesses as small multiples [1] in (3), and variables as a table in (4). *Providers* were already able to change descriptions and variable attributes (*P-T1*).

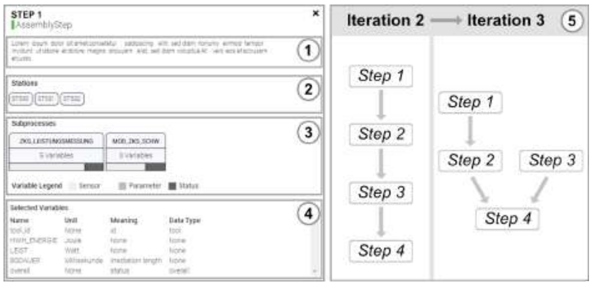


Fig. 7. Tooltip of a selected step on the left and evolution of how we handled step sequences. (1) shows the description of a step, (2) its stations, (3) its subprocesses, and (4) its variables. (5) shows how we changed the representations of step sequences from iteration 2 to 3.

However, *consumers* and *providers* reported that displaying steps as a sequence in one line was not correct, since this did not reflect the real situation of a manufacturing process. As shown in Figure 7 (5), in real world manufacturing processes steps can be root elements of a sequence. For example, step 1 can have step 2 as a successor, but step 3 can be the beginning of a new manufacturing sequence. Thus, step 3 does not follow step 2 but step 2 and step 3 both result in step 4. Furthermore, *consumers* and *providers* reported that they found it difficult to identify the relationship between different nodes and wished to separate parts, steps, and stations, which is why we decided to provide different views for each entity.

Iteration 3: In the last iteration, we separated parts, steps, and stations for an assembly line in separated views as shown in Figure 8 (3). Even though we included variable tables in the previous iteration, especially *consumers* reported that an exemplary distribution of sensor variable recordings would significantly ease their interpretation (*C-T3*). Hence, we visualized variable distributions in a separate view, which represents the lowest possible level of detail. Furthermore, we included interacting mechanisms to connect entities (*P-T2*), such as assigning variables to parts, or connecting stations.

8 Case Study

We provide examples of how our approach supports *providers* and *consumers* with a case study. Case studies are a method to dive deep into a specific domain to provide insights into a phenomenon within its environment [13,41], especially when the practical context and the phenomenon are not clearly evident [55]. Considering the problem and the users, we are interested in how ManKnowVis helps *providers* in externalizing their knowledge and how *consumers* can leverage this previously inaccessible knowledge to better perform use cases. Since these processes are difficult to observe outside of organizational context, we consider a case study as an appropriate approach [21].

Participants: We performed the case study for an assembly line for battery modules in collaboration with seven participants (two female, five male). For the *providers*, we recruited one *planner*, with general knowledge about the entire assembly line and two *technologists* responsible for specific manufacturing stations in the same assembly line. For the *consumers*, we recruited three *implementers* involved in use cases from the assembly line and one *assessor*, who evaluated use cases about the assembly line. The study participants were between 28 and 40 years of age.

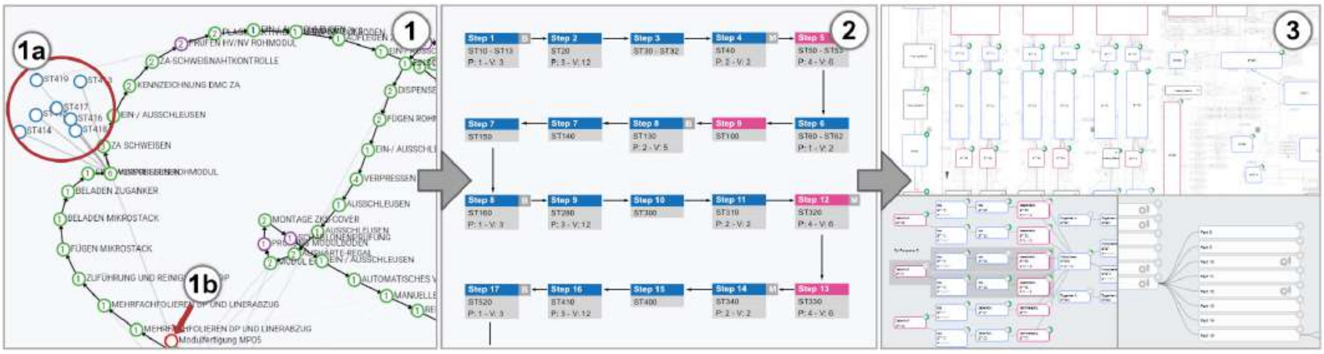


Fig. 8. Evolution of step and stations sequences. In a first attempt in (1), we used a force layout, which we changed to a directed sequence of nodes containing steps and stations in (2). In (3), we display parts, steps, and stations in individual views.

Data: For the case study, we uploaded data from all six described data sources of Section 4.3. In this regard, a battery module consisted of over 15 parts (e.g., heat-conducting sheet, battery cell). For eight parts, CAD drawings were available. Parts were produced at over 50 steps, manufactured at over 100 stations, and contained over 1000 variables. Furthermore, we uploaded sensor measurements from 753 randomly selected battery modules.

Procedure: First, we performed kick-off meetings with all study participants to introduce all features of ManKnowVis. Sessions were held online and took from 60 - 90 minutes. After that participants used ManKnowVis for two weeks. In this regard, we asked *providers* to document entities in the form of describing and connecting entities. *Consumers* were encouraged to analyze entities and evaluate whether they were able to use available information for use cases. The use cases were about identifying whether cell contacting systems were correctly attached to battery modules and the evaluation of cell thicknesses for their compression into modules. After the study, we interviewed *providers* and *consumers* to evaluate how ManKnowVis supported them with their tasks.

8.1 Results for Providers

Providers reported that they were using ManKnowVis for 90 minutes on average during the case study. Furthermore, *providers* described how ManKnowVis supported them in their documentation efforts. Especially the feature of connecting variables to parts was noted as helpful where one *provider* noted that *"I cannot remember how often we executed the exact same annotation task, where I drew rectangles of CAD images and saved them on my local file storage"*. Furthermore, *providers* reported that documenting knowledge with ManKnowVis would impact use cases positively where one *provider* made the following example: *"I am currently involved in an ongoing data analysis project with two technologists and another data scientist. In this project, we have to do the same analysis for a large data set every three months. However, every time we have to do this analysis, we somewhat have to start from the beginning because we have to figure out which exact part we have to analyze, which variables are recorded, and how the variables are related to the part. Using ManKnowVis would have saved us a lot of time in the past because we could have documented exactly this information"*. Finally, *providers* were satisfied with the accuracy of how we managed to automatically position stations on the shop floor image using computer vision since only 2% of all stations had the wrong position.

Providers also recommend some system improvements, which we will include in further iterations of ManKnowVis. First, they requested to include more data sources for entities to improve their own analyses. For example, process flow diagrams show how parts circulate in an assembly line. Here, *providers* suggested tracking individual parts along with their stations in the manufacturing process. Another suggestion was to visualize multiple floors of an assembly line. For instance, sometimes parts are transferred to elevators and then continue on the ceiling of the shop floor. Even though our analyzed assembly line did not include elevators it is necessary to visualize such processes for additional assembly lines.

8.2 Results for Consumers

Consumers reported that ManKnowVis helped them in better comprehending use cases. For instance, ManKnowVis supported them in understanding what data is currently available for the assembly line, where one *consumer* noted that "*seeing which variables for stations are available helps us in identifying which data we can use for a use case"*. Furthermore, ManKnowVis provided them with a good overall understanding of the assembly line. In this case, a *consumer* explained that *"since the station view is very similar to what planners are using it feels a little bit like we see the world through a planner's eye. This helps us in understanding how they work and what they need from us during use cases."* Also, metadata attributes of variables, such as units were noted as helpful to understand the meaning of variables. Finally, we outlined three situations in Figure 9, where *consumers* showed us how ManKnowVis helped them in better understanding variables for a use case. Figure 9 (1) shows how coloring different stations impacts the analysis of two variables. Before ManKnowVis, *consumers* did not make a differentiation between stations. However, in some cases, this differentiation is necessary, since due to distinct calibrations of stations, stations sometimes measure variables slightly differently. Furthermore, ManKnowVis supported *consumers* in identifying variables, which contain outliers (see Figure 9 (2)). Here one *consumers* mentioned that *"some variables contained some outliers, which I will have a closer look at."* ManKnowVis also provides information about which variables might be excluded from further analyses. For example, Figure 9 (3) illustrates one variable which always records the same value. Such variables might be of less importance for use cases.

As system improvements, *consumers* suggested including a new view to visualize knowledge maturity levels of entities as ordered list to immediately see, which entities need enhancements. Furthermore, they wished to include more assembly lines to evaluate how parts, steps, and stations are related in a broader context.

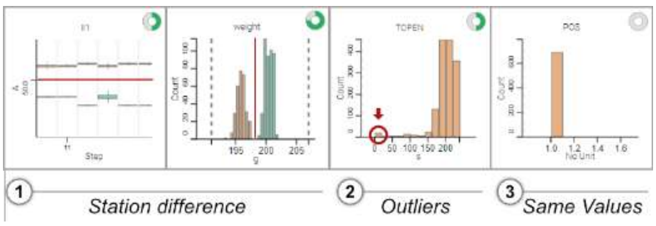


Fig. 9. Three different scenarios of how ManKnowVis supported consumers in analyzing variables. (1) shows two variables that must be analyzed separately, since their stations perform slightly different measurements. (2) illustrates variables that contain outliers and (3) variables that record the same value being less adequate for deeper analyses.

9 Discussion

In this design study, we contribute the problem characterization and abstraction of the externalization and contextualization of knowledge about a manufacturing process for battery modules. We further report the interactive design of the presented knowledge-assisted visualization system ManKnowVis, which we evaluated together with seven *consumers* and *providers*. Our design study represents a very detailed view of the problem domain. From a more abstract point of view, we integrated a knowledge-assisted visualization system into the workflows of *providers* to support their documentation efforts and of *consumers* to help them in understanding complex manufacturing data to better perform use cases. To the best of our knowledge, the problem we described in Section 4.2 is unique inside BMW and has not been addressed by other researchers so far. We thus argue that a design study with a detailed analysis of the problem domain and resulting abstractions was necessary to successfully support related stakeholder groups with their very specific tasks.

The overall usefulness of our approach is demonstrated by the fact that both *providers* and *consumers* reported that they were able to reach their goals outlined in Section 5.2 better with ManKnowVis. Before ManKnowVis, *providers* had to externalize their knowledge mostly manually in local documents or in personal discussions with *consumers*. In turn, *consumers* had no system at hand, that contextualizes knowledge from different databases across BMW. Our evaluation, however, is constrained by the fairly small sample size of participants. However, it is rather unusual to build solutions for specific domain problems, where more than a couple of experts are actually able to evaluate the proposed approaches as demonstrated by other case study evaluations [2, 4, 16, 23, 47]. Nevertheless, we believe that our system is a step towards bridging the gap between *providers* and *consumers* and allows *consumers* to access previously inaccessible knowledge to better perform use cases.

Although the design was specific to knowledge about one assembly line at BMW, there are some aspects that provide guidance to the design in other manufacturing related domains. Some of that guidance results in the following suggestions to contextualize knowledge for complex domains and design similar systems:

1) Use Ontologies to Abstract from Data: The ontology guided us in connecting different existing data sources. For example, how stations are related to steps. All data sources are currently available at different databases at BMW but before ManKnowVis, *consumers* had to analyze documents in each database individually. Ontologies also support the definitions of knowledge maturity levels. In our case, we were able to derive the exact information necessary to complete the documentation of entities and their relations, such as stations that need a sub-line. We therefore were able to directly address knowledge maturity levels with the specific *provider* tasks, *describe* and *connect* entities. The resulting guided workflows supported *providers* in easily enhancing all used data sources.

2) Consider Different *Providers* and *Consumers*: As outlined in Section 5.2, both *providers* and *consumers* have significantly different goals and needs, which must be addressed separately. One might argue, that each user group deserves an individual system, for instance, a documentation system for *providers* and a system showing relations between existing data for *consumers*. However, we believe that both users groups should not be separated strictly from each other. For instance, *technologists* might in one situation take the role of *providers* in describing a station but in another situation also be *consumers* analyzing other stations they are not responsible for. This situation was also reflected in our study, where one *provider* noted, that *"sometimes we also need to know something about steps or station we are not responsible for."* Therefore, we argue that in our case it was necessary to build a system that addresses the needs of two separate user groups.

3) Use Linked Views for Individual Entities: In a first attempt, we tried to build a system showing all entities and their relations in a single force-directed layout. However, as outlined in Section 7, our study participants reported that this did not reflect the real process of an assembly line. Hence, we propose to represent entities with simple individual views, such as tree layouts or process diagrams. Before, users were not able to inspect an assembly line with multiple level of detail views. Now, users can gather information about the manufacturing process using different entry points into analyses. For example, they could choose a top-down approach inspecting first a part and all its related steps, or a bottom-up approach starting at one station and evaluating, which other stations are related to it. Before that, especially *consumers* had no system at hand to perform such analyses.

4) Consider Annotations to Externalize Knowledge: We present guided workflows with our visualization on how to externalize and store knowledge. Besides simple features, such as text inputs for variable descriptions, we included more abstract functionalities in the form of annotations. Here, drawing sub-lines allowed to cluster stations, while drawing rectangles on CADs allowed *providers* to connect variables to specific part regions. This form of externalizing knowledge is currently completely new at BMW. It does not only help *providers* in easily performing their documentation tasks but also *consumers* involved in current use cases or even outside the application domain [11]. Annotations of variables, for instance, can be queried by external stakeholders, such as suppliers in charge of developing specific manufacturing stations to evaluate how variables are related to specific regions of a part.

We believe that these four recommendations can help other researchers when investigating similar manufacturing domains. For example, Sun et al. [47] developed the system *PlanningVis*, which also visualizes part dependencies to support the exploration and comparison of manufacturing processes. To consider the relation of parts to other entities, they could connect them via ontologies to distinct data sources that provide details about stations or variables.

The design also had some limitations, which we briefly summarize here. In our visualization, we use data from only one assembly line. This is because the current ontology is designed for single assembly lines. A solution for this can be the adaption of the ontology via introducing successor assembly lines, as it is already the case for parts, steps, and stations. To be consistent with our recommendations of providing views for separate entities, this results in the need to introduce an additional view to also navigate between assembly lines. Furthermore, we do not consider the fact that assembly lines can have different floors to transport parts. A solution for that could be to either include different filters in the station view to show different floors or to represent the assembly line in a three dimensional space.

10 Conclusion and Future Work

This paper presents a design study on the development of a knowledge-assisted visualization approach to contextualize and leverage the knowledge of a manufacturing process for battery modules. The resulting system ManKnowVis addresses two different user groups with distinct levels of manufacturing knowledge. *Providers* can use ManKnowVis to externalize their knowledge via describing and connecting different entities, such as parts, steps, stations, or variables. They can do so with different techniques, such as providing text inputs or annotating entities. *Consumers* can use ManKnowVis to access previously inaccessible knowledge to better understand data-driven manufacturing use cases. ManKnowVis comprises three different level of detail views. (1) The parts and steps view gives an overview of parts as a tree and steps as Sugiyama layout, where users can drill down to details of parts via inspecting CAD drawings or steps via table views. (2) As a middle level of detail, the station view displays stations as rectangles on top of the real shop floor layout. (3) The variable view provides the highest level of detail and shows variable metadata attributes and their distributions. ManKnowVis evolved iteratively in close cooperation with *providers* and *consumers* from BMW. The success of our design is shown in a case study, where we outline different success scenarios how it helped in externalizing and comprehending complex manufacturing knowledge.

There are several avenues for our research. One is to investigate the flow of parts through an assembly line to identify bottlenecks or to trace back errors and the resulting challenges for future visualizations. A second challenge is to evaluate how additional manufacturing entities, such as plants, can be added to the visualization interfaces of ManKnowVis. Finally, the generalizability of our visualization concept should be investigated in other application domains.

References

[1] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small multipiles: Piling time to explore temporal patterns in dynamic networks. *Computer Graphics Forum*, 34, 2015.

[2] A. Barsky, T. Munzner, and R. Kincaid. Cerebral: Visualizing multiple experimental conditions on a graph with biological context. *IEEE transactions on visualization and computer graphics*, 14:1253–60, 11 2008.

[3] S. Berg, D. Kutra, T. Kroeger, C. Straehle, B. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Köthe, F. Hamprecht, and A. Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16:1–7, 2019.

[4] M. Booshehrian, T. Möller, R. Peterman, and T. Munzner. Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making. *Computer Graphics Forum*, 31:1235–1244, 06 2012.

[5] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1):12–19, 2009. doi: 10.1109/MCG.2009.6

[6] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. Paved: Pareto front visualization for engineering design. *Computer Graphics Forum*, 39:405–416, 2020.

[7] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek. A brief introduction to opencv. In *2012 Proceedings of the 35th International Convention MIPRO*, pp. 1725–1730, 2012.

[8] M. Eiglsperger, M. Siebenhaller, and M. Kaufmann. An efficient implementation of sugiyama's algorithm for layered graph drawing. In J. Pach, ed., *Graph Drawing*, pp. 155–166. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[9] J. Eirich, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. *Transactions on Visualization and Computer Graphics*, 07 2021.

[10] J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *Visualization in Data Science at IEEE VIS*, 2020.

[11] J. Eirich, D. Jäckle, S. Werrlich, and T. Schreck. Visual analytics in organizational knowledge creation: A case study. In *European Conference on Information Systems*, 04 2021.

[12] P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 92–103, 2017. doi: 10.1109/VAST.2017.8585498

[13] B. Flyvbjerg. Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2):219–245, 2006.

[14] T. Fujiwara, S. Chandrasegaran, M. Brundage, T. Sexton, A. Dima, and K. Ma. A visual analytics approach for the diagnosis of heterogeneous and multidimensional machine maintenance data. In *Proceedings - 2021 IEEE 14th Pacific Visualization Symposium, PacificVis 2021*, IEEE Pacific Visualization Symposium, pp. 196–205. IEEE, United States, 2021.

[15] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From visual exploration to storytelling and back again. *Computer Graphics Forum (EuroVis '16)*, 35(3):491–500, jun 2016.

[16] H. Guo, S. R. Gomez, C. Ziemkiewicz, and D. H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):51–60, jan 2016.

[17] F. Heimerl, S. Koch, H. Bosch, and T. Ertl. Visual classifier training for text document retrieval. *Transactions on Visualization and Computer Graphics*, 18(12):2839–2848, 2012.

[18] B. J. Kadlec, H. M. Tufo, and G. A. Dorn. Knowledge-assisted visualization and segmentation of geologic features. *IEEE Computer Graphics and Applications*, 30(1):30–39, 2010.

[19] A. Kampker, K. Kreiskother, N. Lutz, V. Gauckler, and M. Hehl. Re-ramp-up management of scalable production systems in the automotive industry. In *Industrial Technology and Management*, pp. 137–141, 2019.

[20] A. Kampker, A. Maue, C. Deutskens, and R. Forstmann. Standardization and innovation: Dissolving the contradiction with modular production architectures. *2014 4th International Electric Drives Production Conference (EDPC)*, pp. 1–6, 2014.

[21] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.

[22] T. Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. Wijk, J.-D. Fekete, and D. Fellner. Visual analysis of large graphs: State-of-the-art and future research challenges. *Comput. Graph. Forum*, 30:1719–1749, 09 2011.

[23] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2498–2507, dec 2011.

[24] A.-P. Lohfink, S. D. Duque Anton, H. Leitte, and C. Garth. Knowledge rocks: Adding knowledge assistance to visualization systems. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1117–1127, 2022.

[25] B. Lorensen. On the death of visualization. In *Position Papers NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges*, vol. 1, p. 5, 2004.

[26] G. Machado, M. Oliveira, and L. Fernandes. A physiologically-based model for simulation of color vision deficiency (vol 15, pg 1291, 2009). *IEEE transactions on visualization and computer graphics*, 15:1291–8, 11 2009.

[27] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20:1633–1642, 12 2014.

[28] S. Mazumdar, A. VARGA, V. Lanfranchi, D. Petrelli, and F. Ciravegna. A knowledge dashboard for manufacturing industries. In *The Semantic Web: ESWC 2011 Workshops*, vol. 748, pp. 112–124, 05 2011.

[29] S. Miksch, H. Leitte, and M. Chen. *Knowledge-Assisted Visualization and Guidance*, pp. 61–85. Springer International Publishing, 08 2020.

[30] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):438–448, 2019.

[31] T. Munzner. A nested model for visualization design and validation. *Visualization and Computer Graphics*, 15:921 – 928, 2010.

[32] K. Nie, P. Baltzer, B. Preim, and G. Mistelbauer. Knowledge-Assisted Comparative Assessment of Breast Cancer using Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Computer Graphics Forum*, 2020.

[33] I. Nonaka and H. Takeuchi. *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford University Press, New York, 1995.

[34] M. K. Rahman, M. H. Sujon, and A. Azad. Force2Vec: Parallel Force-Directed Graph Embedding. In *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 442–451. IEEE, Sorrento, Italy, Nov. 2020.

[35] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 230–239, 2017.

[36] T. Ruppert, M. Staab, A. Bannach, H. Lücke-Tieke, J. Bernard, A. Kuijper, and J. Kohlhammer. Visual interactive creation and validation of text clustering workflows to explore document collections. *Electronic Imaging*, 2017(1):46–57, 2017.

[37] R. Sahann, T. Möller, and J. Schmidt. Histogram binning revisited with a focus on human perception. In *2021 IEEE Visualization Conference, IEEE VIS 2021 - Short Papers, New Orleans, LA, USA, October 24-29, 2021*, pp. 66–70. IEEE, 2021.

[38] M. Sedlmair, C. Bernhold, D. Herrscher, S. Boring, and A. Butz. Mostvis: An interactive visualization supporting automotive engineers in most catalog exploration. In *2009 13th International Conference Information Visualisation*, pp. 173–182, 2009. doi: 10.1109/IV.2009.95

[39] M. Sedlmair, A. Frank, T. Munzner, and A. Butz. Relex: Visualization for actively changing overlay network specifications. *Visualization and Computer Graphics*, 18:2729–2738, 2012.

[40] M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz. Cardiogram: Visual analytics for automotive engineers. In *Human Factors in Computing Systems*, pp. 1727–1736, 2011.

[41] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics*, 18:2431–2440, 2012.

[42] B. Shneiderman. The eyes have it: A task by data type taxonomy for

information visualizations. In *Visual Languages*, p. 336. IEEE Computer Society, USA, 1996.

[43] R. Smith. An overview of the tesseract ocr engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 629–633, 2007. doi: 10.1109/ICDAR.2007.4376991

[44] T. Sobral, T. Galvão, and J. Borges. An ontology-based approach to knowledge-assisted integration and visualization of urban mobility data. *Expert Systems with Applications*, 150:113260, 2020.

[45] H. Stitz, S. Gratzl, H. Piringer, T. Zichner, and M. Streit. Knowledgepearls: Provenance-based visualization retrieval. *IEEE Transactions on Visualization and Computer Graphics (VAST '18)*, 25(1):120–130, 2018. doi: 10.1109/TVCG.2018.2865024

[46] A. Suh, M. Hajij, B. Wang, C. Scheidegger, and P. Rosen. Persistent homology guided force-directed graph layouts. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):697–707, 2020.

[47] D. Sun, R. Huang, Y. Chen, Y. Wang, J. Zeng, M. Yuan, T.-C. Pong, and H. Qu. Planningvis: A visual analytics approach to production planning in smart factories. *IEEE transactions on visualization and computer graphics*, PP, 08 2019.

[48] J. Suschnigg, B. Mutlu, A. Fuchs, V. Sabol, S. Thalmann, and T. Schreck. Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Big Data Visual Exploration and Analytics*, 2020.

[49] M. Waldner, T. Geymayer, D. Schmalstieg, and M. Sedlmair. Linking unstructured evidence to structured observations. *Information Visualization*, 20(1):47–65, 2021.

[50] C. Wang and K.-L. Ma. Information and knowledge assisted analysis and visualization of large-scale data. In *2008 Workshop on Ultrascale Visualization*, pp. 1–8, 2008. doi: 10.1109/ULTRAVIS.2008.5154057

[51] X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang. Defining and applying knowledge conversion processes to a visual analytics system. *Comput. Graph.*, 33:616–623, 2009.

[52] D. Wu, X. Wang, J. Su, B. Tang, and S. Wu. A labeling method for financial time series prediction based on trends. *Entropy*, 22(10), 2020.

[53] L. Xiao, J. Gerth, and P. Hanrahan. Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation. In *2006 IEEE Symposium On Visual Analytics And Technology*, pp. 107–114. IEEE, Baltimore, MD, USA, Oct. 2006.

[54] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016.

[55] R. K. Yin. *Case Study Research: Design and Methods (Applied Social Research Methods)*. Sage Publications, 2008.

[56] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Transactions on Visualization and Computer Graphics*, PP, 08 2017.

[57] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22:419–435, 2019.

6.6 P6: Visual Analytics in Organizational Knowledge Creation: A Case Study

Full reference: *J. Eirich, D. Jäckle, S. Werrlich, and T. Schreck. Visual analytics in organizational knowledge creation: A case study. In European Conference on Information Systems, 04 2021*

VISUAL ANALYTICS IN ORGANIZATIONAL KNOWLEDGE CREATION: A CASE STUDY

Research paper

Eirich, Joscha, University of Bamberg, Bamberg, Germany, joscha.eirich@uni.bamberg.de

Jäckle, Dominik, BMW Group, Munich, Germany, dominik.jaeckle@bmw.de

Werrlich, Stefan, BMW Group, Munich, Germany, stefan.werrlich@bmw.de

Schreck, Tobias, Graz University of Technology, Graz, Austria, tobias.schreck@cgv.tugraz.at

Abstract

The conversion between tacit and explicit knowledge remains an often-discussed and highly relevant topic in organizational knowledge creation. Although prior research addresses this process, it primarily focuses on the conversion between tacit and explicit knowledge through social processes. This work discusses theories of organizational knowledge creation in the light of sociotechnical systems, and specifically extends them to the interaction between individuals and visual analytics systems that afford analytical decision making based on interactive visualization and knowledge discovery mechanisms. Based on related work, we develop a theoretical framework to explain novel mechanism for knowledge creation afforded by visual analytics systems. We evaluate our framework with a case study with one of the leading organizations in the automotive industry. Over the course of the case study, we observe and analyze interactions between domain experts and a newly introduced visual analytics system. Through our case study findings, we reveal novel mechanisms of organizational knowledge creation and discuss their implications.

Keywords: Knowledge Management, Knowledge Creation, Knowledge Conversion, Visual Analytics, Human-Computer Interaction.

1 Introduction

Knowledge is a primary resource for achieving and maintaining a competitive advantage (Nonaka et al. 2000), leaving knowledge creation as an important task for companies. The process itself is challenging, as it requires accessing the knowledge individuals possess, where they are often not aware of their own tacit or explicit knowledge. Tacit knowledge is bound to the individual and difficult to communicate in contrast to explicit knowledge that can be readily accessed (Nonaka and Takeuchi 1995). Nonaka and Takeuchi (1995) describe knowledge creation as an iterative process in which knowledge is created through the conversion between tacit and explicit knowledge. In this regard, traditional information system technologies have been studied to great extent. However, the role of more recent technologies, specifically those based on artificial intelligence (AI), remain understudied in the context of knowledge creation and conversion (Sanzogni et al. 2017). Hence, we consider the effect of a specific type of systems, visual analytics (VA) systems, on knowledge creation and their affordances to support the conversion between tacit and explicit knowledge. The term affordances is used here to describe possible goal-oriented actions that emerge from the relationship between IT artifacts and specified groups of actors in specific contexts (Hutchby 2001; Vyas et al. 2003; Zammuto et al. 2007).

VA is a branch of information visualization focused on the interactive visualization of information. It encompasses various aspects of analytical tasks in decision-making, including the pre-processing of raw data, building models and deriving recommendations from them, as well as their visualization to facilitate the generation of new knowledge (Thomas and Cook 2005). VA systems draw upon methods and techniques from information visualization and AI to afford their users the ability to draw conclusions based on available data, models, and recommendations, and to facilitate sensemaking (Keim 2010). Providing these affordances, VA systems play a focal role in knowledge creation (Federico et al. 2017). Research on the intersection between VA and knowledge creation has recognized their relevance (Fayyad et al. 1996; Sacha et al. 2014; van Wijk 2005) but has paid little attention to the process of knowledge conversion, especially in collaborative settings that go beyond individual interactions with a VA system. We address this gap with the following research question:

“*What novel mechanisms of organizational knowledge creation and conversion are afforded by visual analytics systems?*”

In answering this question, we aim to extend the work of others who have explored aspects of knowledge creation in the context of VA (Keim et al. 2008; Sacha et al. 2014). We aim to achieve this goal by drawing on the work of Nonaka and Takeuchi (1995) on the process of knowledge creation and conversion. Based on related work, we derive a theoretical framework and evaluate it empirically through a case study of a company in the automotive industry. Following this research design, we want to understand and conceptualize the process of knowledge creation and conversion in the presence of a VA system that proactively augments decision-making and cognitive reasoning. Hence, our contributions are as follows: *(1) a framework, explaining the process of organizational knowledge creation and conversion afforded by VA systems;* and *(2) provision of novel mechanisms for organizational knowledge creation afforded through VA.*

2 Theoretical Background

The following are descriptions of well-established models for knowledge creation in the context of organizations and VA. First, we describe the processes of organizational knowledge creation in Section 2.1. Next, we give an overview of popular models of knowledge creation in VA in Section 2.2.

2.1 Organizational Knowledge Creation

According to the knowledge-based view, knowledge creation capabilities are a strategic asset that contribute to organizations by helping to improve organizational performance (Grant 1996). Nonaka and Takeuchi (1995) contributed to the area of knowledge creation with their theory of knowledge creation, dividing knowledge into the two types: *explicit* knowledge that can be externalized easily, for example, in words and numbers; and *tacit* knowledge that is inherent to the individual. The latter often not recognized by the individual as knowledge but rather is expressed through action, commitment, and involvement, which renders it notoriously difficult to codify and externalize (Wang et al. 2009).

Nonaka and Takeuchi proposed their SECI model to conceptualize the process of knowledge creation and the conversion from one knowledge type to another. It comprises the processes of socialization (S), externalization (E), combination (C), and internalization (I). During interactions between individuals (*socialization*), tacit knowledge is converted through shared mental models. Resulting tacit knowledge is converted into explicit knowledge by codifying it (e.g., in the form of words; *externalization)*. Externalized sources of explicit knowledge can be combined into more systematic and comprehensive sets of explicit knowledge (*combination)*. Further, explicit knowledge can be transformed into tacit knowledge by individuals, which is closely related to the concept of "learning by doing" (Nonaka et al. 2000; *internalization*). The sequential iteration of each process phase of the SECI model creates a spiral of new knowledge. To increase knowledge, a central task of organizations is to ensure knowledge creation and conversion. (Iyengar et al. 2015).

Prior research supports the notion that the use of IT is positively related to the collection, storage, and dissemination of knowledge in organizations (Argote and Miron-Spektor 2011; Iyengar et al. 2015; Kim et al. 2016; Kyriakou et al. 2017; Trantopoulos et al. 2017). This previous work, however, focuses on storing and sharing information (Kim et al. 2016; Kyriakou et al. 2017; Trantopoulos et al. 2017), neglecting the role of model-based reasoning or interactive data visualization in knowledge creation in collaborative contexts, as they are afforded by VA systems (Nobarany et al. 2012). We build on this prior research to include these additional dimensions and analyze their effects on knowledge creation between VA users and their broader environment.

2.2 Knowledge Creation through Visual Analytics

VA has been described as "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2005). By using methods from knowledge discovery in databases, statistics, and AI, VA systems provide a range of affordances to perform analyses. The systems suggest relevant insights and analytical interfaces to human users, who in turn review and revise the systems' output (Keim et al. 2008). VA is closely related to the discipline of *business intelligence* (BI). In BI, data are also gathered, analyzed, and transformed into information, which can then be converted into new information or knowledge (Shollo and Galliers 2016). While VA and BI overlap, they differ in important aspects. VA is concerned mostly with the representation of complex and often unstructured large data sets, for example, in the context of exploratory malware (Wagner et al. 2017) or speech (Sacha et al. 2018) analyses, and the design of resulting visualization interfaces. In turn, traditional BI tools are developed for gathering and storing small amounts structured data, for example, through relational databases (Shollo and Galliers 2016). Furthermore, the scope of BI is not about the design of specific interfaces for exploratory analyses, but comprises technologies and strategies for the collection and analysis of data to make better-informed decisions (Davenport 2010). VA includes large amounts of both structured and unstructured data and is concerned with specific implementations of systems for exploratory analyses, which is why our focus here is on VA rather than BI.

In VA, a variety of models exist that can be related to knowledge creation, such as the classical visualization pipeline (Card et al. 1999; Card and Mackinlay 1997), the process of knowledge discovery in databases (Fayyad et al. 1996), and the sensemaking loop (Pirolli and Card 2005). The models of van Wijk (2005), who focuses on the system perspective, and Sacha et al. (2014), who emphasize the human perspective, have been widely adopted and enhanced and thus stand out within the literature.

Van Wijk's (2005) operational model of visualization identifies three spaces – the *data*, *visualization* and *user* spaces – to describe the context in which visualizations operate. Wang et al. (2009) extend van Wijk's model by adding a knowledge base to the model, combining different sources of knowledge. They also include the SECI model described earlier. Although, Wang et al. (2009) consider the SECI model, their goal lies in improving visualization systems. Our study goes beyond improving visualization systems and aims at discovering novel mechanisms for organizational knowledge creation through VA. For instance, we find that knowledge within organizations can be externalized in three categories: (1) **Features**: Measurable properties of observed phenomena in analyzed data instances (Bishop 2006), such as the maximum voltage of a battery cell (note that while the term features can also be used to refer to system functionalities, we use the term hereafter only in the context of model inputs); (2) **Labels**: The aggregation of complex information to annotate or label observed data instances, taking prior knowledge into account (Bernard et al. 2018), such as a battery cell that is overheated; and (3) **Models**: Models and rules that describe the behavior of the analyzed data (Green et al. 2009), such as the if-then statement "if the maximum voltage is very high, the possibility of an overheated battery cell increases."

The VA process model of Keim et al. (2010), extended by Sacha et al. (2014), focuses on user operations, combining automated analysis methods with user interactions to gain knowledge. Here, users can choose to apply either *automated* or *visual* analysis methods. When using an *automated* method, data mining techniques are applied to create models that fit original data and reveal interesting properties of the data set. These models are then used to evaluate and improve the visualization. If the user chooses to conduct a *visual analysis*, raw data are mapped directly to specific types of visualizations. Sacha et al. (2014) extend the original Keim et al. (2010) model by describing the process of knowledge creation through the interaction of the user with a VA system. *Figure 1* depicts the Sacha et al. (2014) model; the gray boxes indicate the original model and the white boxes are our extensions. Here, the frequent interaction with the system leads to *findings*. Findings can either result in *actions*, if they are taken as input for additional system interactions, or *insights*, if they are interpreted based on previous domain knowledge. Insights can lead to *hypotheses,* which must be properly tested. To *verify* a hypothesis, a new interaction with the system is triggered by an action.

All the models mentioned describe knowledge creation through the interaction between a human and a VA system, with Wang et al. (2009) even explicitly incorporating the SECI model. Nevertheless, all these models focus on the interaction between a single user and the VA system. We propose that the interaction between humans and machines creates novel mechanisms for organizational knowledge creation (Iyengar et al. 2015). We build on the findings of the studies reviewed here and relate knowledge creation in VA to organizational knowledge creation to derive new, generally applicable mechanisms of knowledge creation afforded through VA.

3 Combining Organizational Knowledge Creation with VA

Both in the context of organizational knowledge management and visual analytics two well-known models, which are the SECI model (Section 2.1) from Nonaka and Takeuchi (1995) and the process of knowledge creation in VA Section 2.2 from Sacha et al. 2014, have been introduced. However, none of the two presented models alone serves to describe the process of organizational knowledge creation and conversion in light of collaborative settings, where more recent technologies, such as VA systems play a focal role (Sanzogni et al. 2017). We propose that novel mechanism for organizational knowledge creation can be explained by combing both the models from Nonaka and Takeuchi (1995) and Sacha et al. (2014).

Since the model of Sacha et al. (2014) focuses primarily on actions performed by the user, we consider it to be the most appropriate model for describing emerging mechanisms for organizational knowledge creation afforded through VA. In turn, the SECI model of Nonaka and Takeuchi (1995) serves well to describe knowledge creation in an organizational context, which emerges above all from the collaboration of individuals.

To build a theoretical framework, which serves to describe novel mechanism afforded through VA systems, we rearranged the existing components from the process of knowledge creation in VA and the SECI model. In the case of the process of knowledge creation in VA, we extend the model to account not only for the process of knowledge creation through direct interactions between users and the system, but also for indirect processes enabled by these interactions that affect knowledge creation within a broader organizational context. To do so, we use the SECI model, where we connect each of the four phases to the model of Sacha et al. (2014).

Our result, shown in *Figure 1*, can be summarized as follows: During frequent interactions with the system, the mental model of the user is adapted continuously in *externalization* and *internalization*, both resulting in an augmentation of the system and the user. Tacit knowledge is *externalized* via the *preparation* of *data*, the *training* of *models* or the *manipulations* of the *visualization* (Section 3.1.2). In turn, explicit knowledge is *internalized* via the *inspection* of the *data*, the *verification* of the *model*, or the *observations* of the *visualization* (Section 3.1.4). While using the system, users can consult colleagues within the organization or share insights with them. This results in new *socialization* cycles triggered via a system interaction (Section 3.1.1). A user's analysis can also result in the reconfiguration and combination of different data instances. This *combination* results from an action the user executes, but is performed by the system by connecting new data instances with each other and providing them to the user through visualizations (Section 3.1.3).

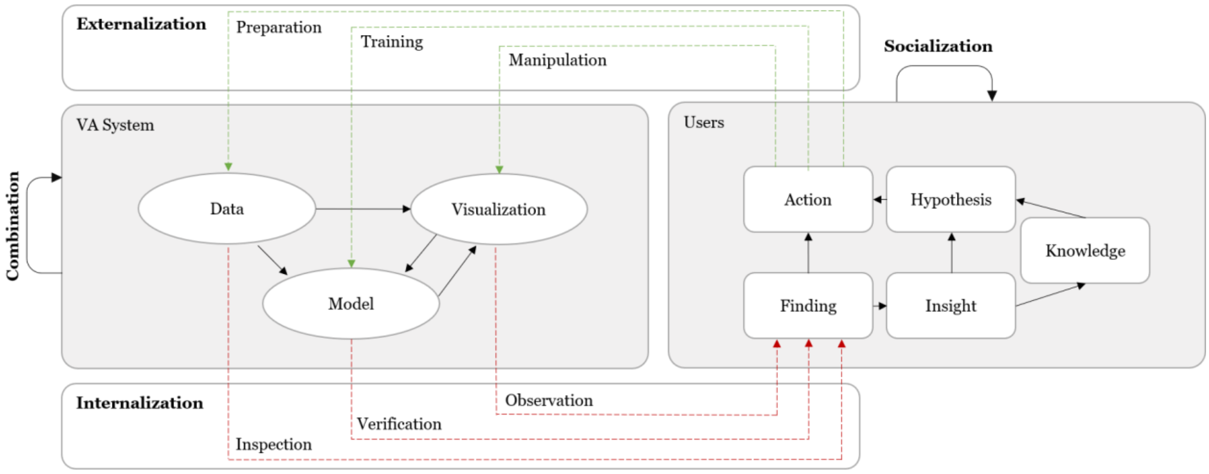


Figure 1: Conversion between tacit and explicit knowledge in VA. Gray boxes indicate components from Sacha et al. (2014), where we also use the SECI phases from Nonaka and Takeuchi (1995).

3.1.1 Socialization

Sanzogni et al. (2017) argue that systems cannot socialize with humans. Thomas and Cook (2005), however, propose that proper visualizations may lead to analytical discourse and the collaboration between system users from system interactions. Thus, even though a user cannot socialize with a system per se, an interaction with the system may trigger socialization with other individuals. Interactions can occur not only between system users, as mentioned by Wang et al. (2009) and Arias-Hernandez et al. (2011) but also with other individuals, such as colleagues from different departments or even outside an organization. For example, a finding as a result of a system interaction (*Figure 1* - *Socialization*), may lead to the inherent wish of a user to consult another individual to validate an assumption. During these interactions, tacit knowledge is transformed through the exchange of metaphors or the adaption of mental models (Nonaka and Takeuchi 1995).

3.1.2 Externalization

Tacit knowledge can be externalized via features, models, or labels during an interaction between humans and machines (Bernard et al. 2018). Externalization in VA systems occurs in a process leading from the user towards the system. Based on context relevant knowledge, the user can interact with the VA system through *manipulation*, *training,* or *preparation* (*Figure 1* - *Externalization*).

Manipulation: As a result of a previously defined goal, manipulations of a visualization can be triggered by user actions (Sacha et al. 2014). Even though the manipulation may not directly result in the augmentation of the system, it may contribute to the generation of new conscious and unconscious findings, which can later be used to train and thereby augment the system (Jackle et al. 2016). Since the visualization is influenced by the recommendations of the underlying model, the model has an indirect effect on actions performed by the user to complete a manipulation task (Green et al. 2009).

Training: After frequent manipulations, a user's understanding of the system can evolve to the extent of being able to train the system consciously. This can happen through the provision either of features or labels. A feature serves as an input for models (Dor and Reich 2012). In the analysis of high-dimensional data sets in particular, expert knowledge about the data is required to select the right features (Thalmann et al. 2018). When users provide labels, they can be employed in combination with features to train a model. The model then interprets defined input features and classifies data according to the patterns derived from the labeled data set with the same features.

Preparation: To improve the system, the user can directly adapt its underlying data representation via data preparation. Here, the data can be prepared via transformations and processing to select features that represent the data (Amershi et al. 2014). The features can be used later as input data for the model, to adapt the visualization, or be stored within an organizational database available to other users.

In addition to externalization as a result of an action, Ragan et al. (2016) summarize different types of provenance during the process of VA. These serve well to externalize knowledge considering the history of changes and advances throughout the analysis process. Considering data and models, the history of data movement and transformation over time and model changes are used to externalize knowledge. For visualization components, the history of graphical views as well as of previous actions and commands serve to reproduce workflows and facilitate other users to understand the system better (Ragan et al. 2016). Thus, knowledge is not only externalized from a single system interaction, but also continuously over the history of system usage.

3.1.3 Combination

Explicit knowledge can be found in different sources, and so the process of combining different types of explicit knowledge is important (Wang et al. 2009). While the combination of different data sources is the result of an action taken by a user, the process of combing data is typically carried out by the system itself (*Figure 1* - *Combination*). Wang et al. (2009) point out that the following must be accounted for to maintain the quality and integrity of combined products of explicit knowledge: the combination of unrelated or incorrect knowledge can degrade the trustworthiness of the system; in addition, it can decrease the quality of the knowledge represented in a visualization. To increase the overall quality and provenance, Federico et al. (2017) suggest increasing collaboration between system users and tracing and verifying relationships between multiple data instances and derived knowledge products.

3.1.4 Internalization

Sacha et al. (2014) argue that experts can generate insights and formulate hypotheses when interacting with a VA system, which can be linked to internalization. Internalization in VA systems occurs in a process leading from the system towards the user. While interacting with the system, users gather relevant evidence to relate this information to prior knowledge, which ideally results in a finding (Endert et al. 2012). This process is closely related to sensemaking (Pirolli and Card 2005) in which, through the interaction with a system, users can explore possible connections between the data, investigate and test hypotheses, and ultimately gain insights. If a finding matches or does not match a previous hypothesis, the mental model of the user changes, which leads to another system interaction or an insight (Chang et al. 2009). In this context, the system interacts with the user by enabling the observation of the visualization, the verification of the model, or the inspection of the data (*Figure 1* - *Internalization*).

Observation: In the visualization of the processed data, units of information are presented by the system in a way that users can perceive them, to focus their attention (Keim et al. 2008). Here, the user is able to discover patterns in the representation of the data (Green et al. 2009). Since the model can change the visualization, observations are also influenced by the reasoning of the model.

Verification: The model aids in collecting relevant evidence to verify a hypothesis that previously was created by the user. The model results can be evaluated by comparing them to data that represent a real-world scenario. If desired, users can define alerts, so the VA system notifies them of changes in the data. Even though the generation of a hypothesis is initiated by the user, the model plays a significant role in neutralizing cognitive biases, shortening the process of analysis, and contributing to the detection of new patterns previously hidden (Green et al. 2009).

Inspection: The interaction with the visualization and the model can lead to a hypothesis that can be evaluated only through interactions with the data. Since it is possible that neither the visualization nor the model represents reality adequately, the user can apply rules directly to the data to collect further evidence that supports a hypothesis (Fayyad et al. 1996).

4 Application of Our Theoretical Framework to the Case of VIMA

Our case study shows the application of the theoretical framework we have described to the project VIMA (Virtual interactive manufacturing assistant). We first describe our case study approach (Section 4.1) before we introduce the context and scope of the VIMA project (Section 4.2) and the data collected (Section 4.3). Finally, we present our findings (Section 4.4) and summarize the implications for organizational knowledge creation (Section 4.5).

4.1 Methodology

We aim to evaluate our theoretical framework (Figure 1) and to use it to reveal and conceptualize novel mechanisms influencing organizational knowledge creation through the frequent interactions between humans and a VA system. Since this process is difficult to observe outside of organizational contexts, we chose a case study approach for an in-depth investigation of the subject matter. A case study is an empirical research approach that enables the analysis of a specific phenomenon within its environment, especially when the boundaries between the practical context and the phenomenon are not clearly evident (Yin 2014). It is useful if the research is not well developed and particularly where the examination of context and dynamics are important (Darke et al. 1998). Case studies, furthermore, can be applied to investigate casual relationships and are a suitable instrument for studying context-rich sociotechnical systems (Yin 2014).

We follow a exploratory single-case study approach (Yin 2014) in which both the phenomenon under investigation and the researcher are assumed to be independent (Sarker et al. 2018). Exploratory case studies are built upon general theories – in our case Nonaka's and Takeuchi's (1995) theory of knowledge creation – that exist to formulate propositions, which are novel mechanisms of knowledge creation in the context of organizational knowledge creation in combination with VA. We chose to carry out a single-case study examining the case of a car manufacturer rolling out a VA system. The chosen case represents a common project among the transformation of many organizations towards a digital company and is suitable for our investigation according to Yin's rational for single-case study designs (Yin 2014). We collected data from different sources that are derived with different methods to capture a wide range of phenomena and processes relevant to our study to ensure its validity (Bonoma 1985).

4.2 Context and Scope

The organization in our case study is a leading automotive manufacturer based in Germany. Like many of its competitors, the manufacturer is investing considerable resources into the development of electric vehicles. Dedicated analytics and engineering teams work collaboratively on establishing automated data-driven manufacturing processes to support this transformation. The units of analysis, in this case, are one department responsible for the manufacturing of *electrical energy storage systems*, one department for *electrical engines*, and a supporting *data analytics* department. We refer to the overall company as AutoCorp, and these units (departments) as AutoCorp Storage (ACS), AutoCorp Engines (ACE), and AutoCorp Analytics (ACA), respectively. AutoCorp faces a quality-control challenge in its manufacturing process. It aims to resolve this challenge by improving the quality of their process in a project spanning all three units.

At present, the components AutoCorp produces are controlled for quality at specialized testing benches, which are optimized to minimize false negatives. During quality assessments, these test stations determine whether tested components meet the quality criteria. If they pass, they are labeled as "OK"; otherwise, they are labeled as "NOK" (not okay). False negative classifications have more severe implications than false positives in producing components for the automotive industry. However, both error types have a negative influence on product quality and cycle times and thus on the costs of the overall manufacturing process.

The project proposed by AutoCorp aims at using models, trained on production data and developed by AutoCorp Analytics, to improve the classification capabilities of the test benches. Building such models requires ACA to access the (tacit and explicit) knowledge of domain experts. To generate and externalize this knowledge, vast amounts of complex high dimensional data sets, comprising machine signals and manually generated annotations of datapoints referring to produced parts, must be analyzed manually by experts. The goal of the VIMA project is to minimize the current error rate of the quality control process (regarding false positives and false negatives) and reduce the additional workload for domain experts from AutoCorp Storage and AutoCorp Engines.

VIMA is a human-in-the-loop VA system that aims to enable domain experts to classify components and generate new knowledge about the underlying business processes. The goal of the system is to visualize huge amounts of high-dimensional, complex machine data, detect anomalies in the manufacturing process, and suggest possible classes for according data instances. VIMA thereby helps with the generation of labels and the discovery of relevant features that can be used to build models. In addition, it supports domain experts in their analysis of parts produced and in generating new knowledge. VIMA was released on March 1, 2020 and is being used by both ACS and ACE. Since its launch, it is continually enhanced by ACA in close cooperation with ACS and ACE.

4.3 Data Collection and Coding

The case study was carried out over a nine-month period. We collected machine sensor-data of AutoCorp's manufacturing process, which underlies the system, conducted semi-structured interviews, and made direct observations. Sensor data comprised 524 recorded measurements of 2,543 electrical engines (98 GB disk space) and 132 measurements of 206 electrical energy storage systems (69 GB of disk space). The sensor data were used as only input for VIMA to enable experts to conduct multiple explorative analyses. Details about VIMA and its input data can be found in (Eirich et al. 2020).

We conducted seven semi-structured interviews with four domain experts for *electrical engines* and three with *electrical energy storage systems* experts. These interviewees were males between ages 24 and 33, had a mean work experience of 9 years, and had worked for AutoCorp for 7 years and 8 months on average. Each expert had a background in mechanical or electrical engineering, and they all indicated a lack of familiarity with data analytics methods. All interviews were held in person and lasted an average of 50 minutes. We questioned interviewees about socialization, externalization, combination, and internalization processes in the context of developing testing procedures for electrical components using VIMA. The interview guidelines comprised 24 questions on the process of knowledge creation during interviewees' daily routines and interactions with VIMA in all stages of knowledge conversion and creation. (Due to space limitations, the full set of questions is not included in this study.) The interviews were recorded, transcribed, and analyzed deductively by the interviewer post hoc. The coding of the interviews was carried out by one researcher according to Hsieh's and Shannon's (2005) method of the *direct content analysis*. In this approach, an existing theory is used as a guide for initial codes to further enhance the theory deductively (Hsieh and Shannon 2005). In our case, we used Nonaka's and Takeuchi's (2000) SECI model and the relations between all four SECI phases to discover novel mechanisms of knowledge creation in the context of organizational knowledge creation in combination with VA.

An on-site researcher made direct observations of the processes being studied and recorded his findings on a weekly basis, following Someren's et al. (1994) Think Aloud Method, through which experts are asked to express their thoughts during their use of a system (Someren et al. 1994). We undertook this with experts from ACS and ACE using VIMA and transcribed the results. We used this additional data source to gain a better understanding of how experts used VIMA and how tacit and explicit knowledge was transformed between system users and VIMA as well as between system and non-system users.

4.4 Findings

At the beginning of the development of VIMA, experts from all three business units (ACS, ACE, and ACA) had to learn the language and technical methods used in the other units. During January 2020, a first prototype was developed by ACA and adapted regularly in close cooperation with experts from ACE and ACS to ensure a shared understanding of VIMA and to enhance the system continually. After this period, the first version of VIMA was deployed by ACA and made available to experts from ACS and ACE. While most use cases carried out with VIMA required substantial in-depth analysis of multiple dependent measurements requiring detailed expert knowledge, we have chosen to show use cases that serve as good examples but will also be easy to understand for readers with no mechanical engineering background. Our results are presented in the following section.

4.4.1 Socialization

Before the launch of VIMA, all the experts regularly sought advice, as part of their daily routines, from other colleagues with in-depth knowledge about specific product details, especially those in R&D departments. After VIMA was introduced and as the experts we studied became increasingly familiar to its functionalities, we observed greater socialization efforts between system users and experts from new domains. The goal of these interactions was to verify assumptions, conduct more detailed inspections of databases, or engage in joint modeling efforts to improve manufacturing processes. As one expert noted, "*The visualization showed me something I did not understand, but after consulting a colleague it helped me to better understand the problem*." Another interesting type of interaction occurred with colleagues from similar domains. Staff of ACE and ACS reported that VIMA provided them with knowledge that increased their bargaining power in meetings or helped them better communicate tacit knowledge. Thus, after the launch of VIMA, socialization was triggered through an interaction with the system, which sometimes resulted in new interactions, which affected a situation not directly related to the initial interaction. For example, experts who interacted with VIMA made new findings and thus often consulted with other experts to discuss these findings; in doing so, new knowledge for experts not associated with VIMA was created. As one domain expert noted, "*At first I did not understand the result of VIMA. However, after calling my colleague, who had a more detailed background of the problem domain I got a better understanding of the initial result*."

4.4.2 Externalization

At the beginning of the VIMA project, experts reported that they externalized tacit knowledge mostly through the documentation of product specifications. Domain experts did this because they wanted to make explicit knowledge available to others at AutoCorp and store knowledge for later use. However, new externalization mechanisms emerged through interaction with VIMA. Rather than dispersing knowledge in the form of documentation, explicit knowledge was stored in the system's database in the form of features, labels, and models. After multiple analyses, experts from both ACS and ACE discovered new, highly relevant features with which to detect specific error types. In addition, the visualizations provided by VIMA revealed previously unknown anomalies in the manufacturing process, where parts could be labeled with corresponding error types. Since these labels were the result of an analysis that could be performed only by experts with tacit knowledge about the part, the labels represent a manifestation of aggregated tacit knowledge. Further, insights derived from the use of VIMA were translated into models that improve the automated assessment and description of product behavior. One domain expert provided an example: "*If the structure-borne noise of our engine increases, the unbalance also increases.*" Just like labels, models denote an aggregation of complex knowledge about product behavior. In contrast to labels, models represent the decision process behind a classification, and thus are a valuable new source of explicit knowledge within AutoCorp.

4.4.3 Combination

Before VIMA, combination of explicit sets of knowledge was primarily carried out by domain experts, working manually. This included combining documents from organizational databases, which often resulted in small sets of explicit knowledge with small and incremental novelty. The combination of high-dimensional sensor data recorded during assembling steps was a time-consuming task and hence not often done by the experts. Thus, resulting data sets contained information only about a few produced parts. VIMA's data-processing pipeline was able to process high-dimensional machine sensor data automatically across the entire manufacturing process, resulting in new comprehensive sets of explicit knowledge. These data sets included information about thousands of produced parts from multiple test stations, which could be filtered by the experts to obtain the most relevant information. In addition, these data sets were used as a source of new inputs to conduct analyses with the system. As one ACE expert said, "*Now, for the first time, we can combine measurements of different sensors across multiple assembling steps.*" Whereas previous data sets had come from single test stations or documents, new explicit knowledge products emerged from the combination of multiple machine sensors across the entire manufacturing process. This did not hold for tacit knowledge, however, which is always bound to individuals.

4.4.4 Internalization

Experts reported that at the beginning of the project they internalized knowledge as a result of incremental learning processes or experiences from previous product development cycles. In addition, they indicated that this process was the result of consciously performing a predefined task, such as reading documentations or consulting a colleague. During the interaction with VIMA, experts from ACS and ACE reported a continual learning process. Throughout these frequent interactions, experts described several findings that, when combined with previous domain knowledge and advice from colleagues, resulted in several insights and "ah-ha" moments (Chang et al., 2009).

Insights were often quite small, as in the case of one ACS expert who found that temperature and voltage, measured during a production process, were not correlated, but independent – in contrast to previously held beliefs. In other cases, insights resulted in formulations of hypotheses, such as "*an increased resistance of a battery module decreases the measured current of an electrical energy storage system.*" Testing these hypotheses created tacit knowledge for the domain experts, since their mental model changed and understanding of the problem increased. This knowledge was quite relevant for the improvement of the system, as it resulted in new valuable features, labels, and improved underlying models for VIMA.

4.5 New Mechanisms for Organizational Knowledge Creation

Based on the application of our theoretical framework to the case study, we derive new mechanisms that should be considered for organizational knowledge creation. They are summarized in *Table 1*.

	Nonaka and Takeuchi (2000)	*New mechanisms for knowledge creation*
Socialization	Knowledge is created as a result of the interaction between individuals and their exchange of tacit knowledge through shared experiences.	Knowledge is created as a result of the interaction between individuals and the system; the interactions can result in knowledge creation between system and non-system users.
Externalization	Externalization is triggered by dialogue or collective reflection. Knowledge is externalized consciously in interaction and articulated via images, symbols, and language.	Externalization is triggered by a system interaction. Knowledge is externalized consciously and unconsciously; it is articulated through features, labels, and models.
Combination	Combination is mostly executed manually and is thus limited to relatively small data sets.	Combination is mostly executed automatically and is thus extended to relatively large data sets.
Internalization	Internalization is triggered by an individual's inherent wish to accumulate organizational know-how. Explicit knowledge must be actualized via action and practice (e.g., reading documents or "learning-by-doing").	Internalization is triggered by the system's notification of changing data instances. Learning is performed by interpreting the system's representation of data and the formulation and evaluation of hypotheses.

Table 1: New mechanisms for organizational knowledge creation from system interactions

During the interaction between individuals, information is interpreted by each individual to become knowledge. VIMA is not able to create knowledge alone; rather, it transforms available data into information and interprets it in a way that creates meaning. Through the interaction of domain experts both from ACS and ACE with the system and each other, we observed the conversion between tacit and explicit knowledge that enhanced the knowledge base of the users and improved the system.

Nonaka and Takeuchi (2000) argue that socialization is a result of direct social interaction; for example, individuals exchange tacit knowledge through hands-on learning experiences. In our case, socialization was triggered through system interactions and often resulted in the creation of new knowledge for individuals using and those not using the system. After multiple analyses with the system, users approached other colleagues to verify assumptions or discuss findings, and VIMA mediated the exchange of tacit knowledge between these individuals. Traditionally, tacit knowledge is externalized consciously through physical interaction, articulated through images, symbols and language. In turn, VIMA facilitates the externalization of tacit knowledge, such as by requiring users to label data. It then combines various labels into larger collections of externalized knowledge, which are then transformed into recommendations and visualizations that are supplied to a wide audience of users. This combination of continuous externalization, automated combination, and visualization affects the internalization of knowledge by VIMA's users. Multiple different sets of explicit knowledge are recombined and reconfigured manually and often contain comparable small data sets. In turn, we observed that combination efforts were mostly carried out by VIMA, with huge data sets recombined into smaller, more comprehensible ones that were easily interpretable by its users. Nonaka and Takeuchi (2000) describe the process of internalizing explicit knowledge as similar to "learning by doing," in which an individual is driven by the inherent wish to learn something new. We observed something different: as experts conducted observations, verifications, and inspections of the data, there was a learning process through which explicit knowledge from data, models, and the visualizations was transformed into new tacit knowledge, such as changed mental models regarding the users' problem domains.

5 Discussion

To gain a better understanding of the process of organizational knowledge creation and conversion in the context of interplay between humans and VA systems, we presented a theoretical framework, which we evaluated with a case study. To address our research question, we outlined traces of new mechanisms for organizational knowledge creation that extend the existing theory of organizational knowledge creation (Nonaka and Takeuchi 1995). Based on our framework (Figure 1), our study results, and resulting derived mechanisms for knowledge creation, we formulate and discuss the following propositions.

1) New tactit knowledge is created through socialization between individuals as a result of system interactions with one user and a VA system, from findings, insights, actions, or hypotheses.

Since the creation of *tacit knowledge* is bound to individuals, VA systems are not capable creating this kind of knowledge. However, VA systems can trigger interactions between multiple individuals through which new tacit knowledge can be created. We observed in our case study that these interactions were resulting from *findings*, *insights*, *actions*, or *hypotheses*. *Actions*, for example, resulted from system interactions in which users made specific observations that resulted in new *findings*. In many instances, these *findings* were discussed with one or multiple other individuals not related to the VA system itself, through which new *insights* or *hypotheses* emerged. After these discussions in particular, the resulting *hypotheses* were tested using proper statistical tests that met the criteria for objectivity, which resulted in new *tacit knowledge* for individuals involved in these analyses. We thus argue that even though *tacit knowledge* is bound to individuals, VA systems trigger *socialization* cycles, which is not considered in the established mechanisms of *socialization* articulated by Nonaka and Takeuchi (2000).

2) Tacit knowledge is converted into explicit knowledge by means of externalization through VA system interactions via manipulating visualizations, training models, or preparing data, and are articulated through features, labels, and models.

Since *tacit knowledge* plays a focal role in the competitiveness of organizations, how organizations can create conditions that enable the *externalization* of *tacit knowledge* (Haldin-Herrgard 2000) is important. We believe that VA systems can contribute significantly in this context and facilitate the conversion from *tacit* to *explicit knowledge* via *visualization manipulations*, *training models*, and *preparing data*. The advantages of VA systems, such as fast filtering of data, adapting interfaces after interactions, and making recommendations based on trained models can be combined with human *tacit knowledge*. The resulting knowledge products *features*, *labels*, and *models*, play an important role in organizational knowledge management and complement well-established but, from our point of view, rather generic explicit knowledge products such as images or symbols (Peltokorpi et al. 2007). Even though *features*, *models*, and *labels* are addressed by many researches (Amershi et al. 2014; Chegini et al. 2020; Jackle et al. 2016; Schneider et al. 2018; Suschnigg et al. 2020), their value for organizational knowledge creation during *externalization* processes has not previously been considered. Thus, organizations should take these knowledge products into account, since they resemble the aggregation of multiple complex sources of *tacit knowledge* (Bernard et al. 2018).

3) New explicit knowledge is created throguh combination abilites of VA systems, in databases, models, or visualizations.

The synthesis of data from multiple sources is an often addressed topic across multiple research communities, such as information technology (He et al. 2010), information systems (Skillicorn and Wang 2001), and pattern recognition (Adhikari and Rao 2008), to name a few. However, to the best of our knowledge, our theoretical framework is the first that relates organizational *combination* efforts of *explicit knowledge* to VA systems. Our case study reveals that an automatic combination of complex multidimensional data was carried out by a VA system, with the results used to trigger learning and interaction cycles. VIMA recombined overwhelming amounts of information into smaller data sets. Thus, we believe that VA systems can create explicit knowledge by providing *models* and *visualizations* to the user and storing resulting knowledge in *databases*. Such knowledge can be created from user interaction with the system by for example adapting models, or automatically, by for example querying databases and presenting recombined results through visualizations. We hence follow the proposition of Wang et al. (2009) that organizations not only should support the synthesis of complex data into cohesive sets of new explicit knowledge but also introduce systems that facilitate the recombination of huge data sets and encourage their validation through experts.

4) Explicit knowledge is converted into tacit knolwedge in internalization as a result of inspecting data, verifying models, or observing visualizations in VA systems.

The process of converting *explicit* knowledge into *tacit* knowledge can best be described as “learning” (Endert et al. 2012; Nonaka et al. 2000). To support learning, VA systems should provide the ability to look at data from different perspectives. This enables users to collect versatile evidence and increase the level of trust in findings or insights (Sacha et al. 2014). In our study, we observed that such learning processes were afforded via system interactions through *inspecting data*, *verifying models*, and *observing visualizations*. In this regard, we agree with Amershi et al. (2014), who stress active user involvement when introducing systems, which can augment cognitive reasoning. In our study, the resulting mutual interaction between humans and machines resulted both in overall continuous system improvement and additional knowledge for its users. Thus, when fostering knowledge creation, organizations should not only consider how to improve the collaboration between their employees but also how to include a system that absorbs *externalized knowledge* over time. The resulting *explicit knowledge* should be incorporated into models that proactively guide users’ decision-making and knowledge creation processes, which are closely related to the notion of provenance in VA by Endert et al. (2012).

We acknowledge that our study has some limitations. The case is dedicated to one company in the automotive industry. Since our study analyzes organizational knowledge creation and conversion afforded by VA systems, our approach is likely to affect the generalizability to knowledge creation and conversion of other context-related systems, such as business intelligence systems. The interpretation of our results is also constrained to our small and biased sample size (i.e., seven male participants). Because VA systems are often designed to tackle very specific domain problems, they naturally address relatively small user groups. Thus, our study results do not allow for drawing any conclusion regarding the empirical validity of our framework beyond our application domain. Even though other described process models (Sections 2.1 and 2.2) address the conversion and creation of knowledge, they take either a system or organizational perspective. We believe that our theoretical framework combines the strengths of both areas and contributes to explaining organizational knowledge creation with VA systems in manufacturing settings.

6 Conclusion and Future Work

In conclusion, we provide novel mechanisms for organizational knowledge creation that are the result of the interaction between humans and a VA system. To derive these mechanisms, we first integrated the process of knowledge creation in VA (Sacha et al. 2014) and the process of organizational knowledge creation and conversion (Nonaka and Takeuchi 1995) into an integrated theoretical framework that accounts for both the interactions between humans and the systems and those between non-system-related individuals. We then conducted out a case study for a 9-month period, collecting data from multiple sources (sensor data, semi-structured interviews, and direct observations). Our findings reveal that the conversion between tacit and explicit knowledge is affected by interactions with VA systems. The resulting novel mechanisms for organizational knowledge creation were clearly noticeable in each phase of the SECI model (Nonaka and Takeuchi 1995). Our derived theoretical framework and novel mechanism for organizational knowledge creation serve as a guiding framework for further research.

Finally, although our findings inform the theory of organizational knowledge creation and conversion, additional research is needed to extend this theory. For example, we focus on a single stakeholder group for our VA system, but research suggests that externalized knowledge in particular can affect other stakeholders within an organization (Markovic and Bagherzadeh 2018). Thus, future research should consider how to diffuse this kind of knowledge to other related stakeholders (e.g., suppliers, analytics departments, or management) and analyze the effect on stakeholders beyond the application domain.

7 References

Adhikari, A., and Rao, P. R. 2008. "Synthesizing Heavy Association Rules from Different Real Data Sources," *Pattern Recognition Letters* (29:1), pp. 59–71.

Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. 2014. "Power to the People: The Role of Humans in Interactive Machine Learning," *AI Magazine* (35:4), p. 105.

Argote, L., and Miron-Spektor, E. 2011. "Organizational Learning: From Experience to Knowledge," *Organization Science* (22:5), pp. 1123–1137.

Bernard, J., Zeppelzauer, M., Lehmann, M., Müller, M., and Sedlmair, M. 2018. "Towards User-Centered Active Learning Algorithms," *Computer Graphics Forum* (37:3), pp. 121–132.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*, Information Science and Statistics, New York: Springer.

Bonoma, T. V. 1985. "Case Research in Marketing: Opportunities, Problems, and a Process," *Journal of Marketing Research* (22:2), p. 199.

Card, S. K., and Mackinlay, J. 1997. "The Structure of the Information Visualization Design Space," in *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, Phoenix, AZ, USA: IEEE Comput. Soc, pp. 92-99,.

Card, S. K., Mackinlay, J. D., and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*, The Morgan Kaufmann Series in Interactive Technologies, San Francisco, Calif: Morgan Kaufmann Publishers.

Chang, R., Ziemkiewicz, C., Green, T. M., and Ribarsky, W. 2009. "Defining Insight for Visual Analytics," *IEEE Computer Graphics and Applications* (29:2), pp. 14–17.

Chegini, M., Bernard, J., Cui, J., Chegini, F., Sourin, A., Andrews, K., and Schreck, T. 2020. "Interactive Visual Labelling versus Active Learning: An Experimental Comparison," *Frontiers of Information Technology & Electronic Engineering* (21:4), pp. 524–535.

Darke, P., Shanks, G., and Broadbent, M. 1998. "Successfully Completing Case Study Research: Combining Rigour, Relevance and Pragmatism," *Information Systems Journal* (8:4), pp. 273–289.

Davenport, T. H. 2010. "Business Intelligence and Organizational Decisions:," *International Journal of Business Intelligence Research* (1:1), pp. 1–12.

Dor, O., and Reich, Y. 2012. "Strengthening Learning Algorithms by Feature Discovery," *Information Sciences* (189), pp. 176–190.

Eirich, J., Jäckle, D., Schreck, T., Bonart, J., Posegga, O., and Fischbach, K. 2020. "VIMA: Modeling and Visualization of High Dimensional Machine Sensor Data Leveraging Multiple Sources of Domain Knowledge," *Symposium on Visualization in Data Science (VDS) at IEEE VIS 2020, 2020*.

Endert, A., Fiaux, P., and North, C. 2012. "Semantic Interaction for Sensemaking: Inferring Analytical Reasoning for Model Steering," *IEEE Transactions on Visualization and Computer Graphics* (18:12), pp. 2879–2888.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. 1996. "From Data Mining to Knowledge Discovery in Databases," *AI Magazine* (17:3), pp. 37–54.

Federico, P., Wagner, M., Rind, A., Amor-Amoros, A., Miksch, S., and Aigner, W. 2017. "The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics," in *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Phoenix, AZ: IEEE, October, pp. 92–103.

Grant, R. M. 1996. "Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration," *Organization Science* (7:4), pp. 375–387.

Green, T. M., Ribarsky, W., and Fisher, B. 2009. "Building and Applying a Human Cognition Model for Visual Analytics," *Information Visualization* (8:1), pp. 1–13.

Haldin-Herrgard, T. 2000. "Difficulties in Diffusion of Tacit Knowledge in Organizations," *Journal of Intellectual Capital* (1:4), pp. 357–365.

He, D., Wu, X., and Zhu, X. 2010. "Rule Synthesizing from Multiple Related Databases," in *Advances in Knowledge Discovery and Data Mining* (Vol. 6119), Lecture Notes in Computer Science, M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 201–213.

Hsieh, H.-F., and Shannon, S. E. 2005. "Three Approaches to Qualitative Content Analysis," *Qualitative Health Research* (15:9), pp. 1277–1288.
Hutchby, I. 2001. "Technologies, Texts and Affordances," *Sociology* (35:2), pp. 441–456.
Iyengar, K., Sweeney, J. R., and Montealegre, R. 2015. "Information Technology Use as a Learning Mechanism: The Impact of IT Use on Knowledge Transfer Effectiveness, Absorptive Capacity, and Franchisee Performance," *MIS Quarterly* (39:3), pp. 615–641.
Jackle, D., Fischer, F., Schreck, T., and Keim, D. A. 2016. "Temporal MDS Plots for Analysis of Multivariate Data," *IEEE Transactions on Visualization and Computer Graphics* (22:1), pp. 141–150.
Keim, D. (ed.). 2010. *Mastering the Information Age: Solving Problems with Visual Analytics*, Goslar: Eurographics Association.
Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H. 2008. "Visual Analytics: Scope and Challenges," in *Visual Data Mining* (Vol. 4404), S. J. Simoff, M. H. Böhlen, and A. Mazeika (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 76–90.
Kim, S. H., Mukhopadhyay, T., and Carnegie Mellon University. 2016. "When Does Repository KMS Use Lift Performance? The Role of Alternative Knowledge Sources and Task Environments," *MIS Quarterly* (40:1), pp. 133–156.
Kyriakou, H., Nickerson, J. V., and Sabnis, G. 2017. "Knowledge Reuse for Customization: Metamodels in an Open Design Community for 3D Printing," *MIS Quarterly* (41:1), pp. 315–332.
Markovic, S., and Bagherzadeh, M. 2018. "How Does Breadth of External Stakeholder Co-Creation Influence Innovation Performance? Analyzing the Mediating Roles of Knowledge Sharing and Product Innovation," *Journal of Business Research* (88), pp. 173–186.
Nobarany, S., Haraty, M., and Fisher, B. 2012. "Facilitating the Reuse Process in Distributed Collaboration: A Distributed Cognition Approach," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, Seattle, Washington, USA: ACM Press, p. 1223.
Nonaka, I., and Takeuchi, H. 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*, New York: Oxford University Press.
Nonaka, I., Toyama, R., and Konno, N. 2000. "SECI, Ba and Leadership: A Unified Model of Dynamic Knowledge Creation," *Long Range Planning* (33:1), pp. 5–34.
Peltokorpi, V., Nonaka, I., and Kodama, M. 2007. "NTT DoCoMo's Launch of I-Mode in the Japanese Mobile Phone Market: A Knowledge Creation Perspective," *Journal of Management Studies* (44:1), pp. 50–72.
Pirolli, P., and Card, S. 2005. *The Sensemaking Process and Leverage Points for Analyst Technology as Identified through Cognitive Task Analysis*.
Ragan, E. D., Endert, A., Sanyal, J., and Chen, J. 2016. "Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes," *IEEE Transactions on Visualization and Computer Graphics* (22:1), pp. 31–40.
Sacha, D., Kraus, M., Bernard, J., Behrisch, M., Schreck, T., Asano, Y., and Keim, D. A. 2018. "SOMFlow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance," *IEEE Transactions on Visualization and Computer Graphics* (24:1), pp. 120–130.
Sacha, D., Stoffel, A., Stoffel, F., Kwon, B. C., Ellis, G., and Keim, D. A. 2014. "Knowledge Generation Model for Visual Analytics," *IEEE Transactions on Visualization and Computer Graphics* (20:12), pp. 1604–1613.
Sanzogni, L., Gustavo, G., and Busch, P. 2017. "Artificial Intelligence and Knowledge Management: Questioning the Tacit Dimension," *Prometheus* (35:1), pp. 37–56.
Sarker, S., Xiao, X., Copenhagen Business School, Beaulieu, T., Utah State University, Lee, A. S., and Virginia Commonwealth University. 2018. "Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 1/2)," *Journal of the Association for Information Systems* (19), pp. 752–774.
Schneider, B., Jackle, D., Stoffel, F., Diehl, A., Fuchs, J., and Keim, D. 2018. "Integrating Data and Model Space in Ensemble Learning by Visual Analytics," *IEEE Transactions on Big Data*, pp. 1–1.
Shollo, A., and Galliers, R. D. 2016. "Towards an Understanding of the Role of Business Intelligence Systems in Organisational Knowing: The Role of BI Systems in Organisational Knowing," *Information Systems Journal* (26:4), pp. 339–367.

Skillicorn, D. B., and Wang, Y. 2001. "Parallel and Sequential Algorithms for Data Mining Using Inductive Logic," *Knowledge and Information Systems* (3:4), pp. 405–421.

Someren, M. W. van, Barnard, Y. F., and Sandberg, J. A. C. 1994. *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*, London: Academic Press.

Suschnigg, J., Mutlu, B., Fuchs, A. K., Sabol, V., Thalmann, S., and Schreck, T. 2020. "Exploration of Anomalies in Cyclic Multivariate Industrial Time Series Data for Condition Monitoring," in *Proceedings of the EDBT/ICDT 2020 Joint Conference Workshops*, CEUR Workshop Proceedings.

Thalmann, S., Mangler, J., Schreck, T., Huemer, C., Streit, M., Pauker, F., Weichhart, G., Schulte, S., Kittl, C., Pollak, C., Vukovic, M., Kappel, G., Gashi, M., Rinderle-Ma, S., Suschnigg, J., Jekic, N., and Lindstaedt, S. 2018. "Data Analytics for Industrial Process Improvement A Vision Paper," in *2018 IEEE 20th Conference on Business Informatics (CBI)*, Vienna: IEEE, July, pp. 92–96.

Thomas, J. J., and Cook, K. A. (eds.). 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, Los Alamitos, Calif: IEEE Computer Soc.

Trantopoulos, K., von Krogh, G., Wallin, M. W., and Woerter, M. 2017. "External Knowledge and Information Technology: Implications for Process Innovation Performance," *MIS Quarterly* (41:1), pp. 287–300.

Wagner, M., Sacha, D., Rind, A., Fischer, F., Luh, R., Schrittwieser, S., Keim, D. A., and Aigner, W. 2017. "Visual Analytics: Foundations and Experiences in Malware Analysis," in *Empirical Research for Software Security* (1st ed.), L. ben Othmane, M. G. Jaatun, and E. Weippl (eds.), CRC Press, pp. 139–171.

Wang, X., Jeong, D. H., Dou, W., Lee, S.-W., Ribarsky, W., and Chang, R. 2009. "Defining and Applying Knowledge Conversion Processes to a Visual Analytics System," *Computers & Graphics* (33:5), pp. 616–623.

van Wijk, J. J. 2005. "The Value of Visualization," in *VIS 05. IEEE Visualization, 2005.*, Minneapolis, MN, USA: IEEE, pp. 79–86.

Yin, R. K. 2014. *Case Study Research: Design and Methods*, (Fifth edition.), Los Angeles: SAGE.

Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., and Faraj, S. 2007. "Information Technology and the Changing Fabric of Organization," *Organization Science* (18:5), pp. 749–762.

6.7 P7: The life cycle of data labels in organizational learning: A Case Study of the Automotive Industry

Full reference: *J. Eirich and F.-P. Diana. The life cycle of data labels in organizational learning: A case study of the automotive industry. In European Conference on Information Systems, 04 2022*

THE LIFE CYCLE OF DATA LABELS IN ORGANIZATIONAL LEARNING: A CASE STUDY OF THE AUTOMOTIVE INDUSTRY

Research Paper

Joscha Eirich, University of Bamberg, Bamberg, Germany, joscha.eirich@uni-bamberg.de

Diana Fischer-Preßler, University of Bamberg, Bamberg, Germany, diana.fischer-pressler@uni-bamberg.de

Abstract

Data labels are an integral input to develop machine learning (ML) models. In complex domains, labels represent the externalized product of complex knowledge. While prior research discussed labels typically as input of ML models, we explore their role in organizational learning (OL). Based on a case study of a German car manufacturer, we contextualize a framework of OL to the use of labels in organizations informing about organizational members who work with labels, requirements of label-based tools, label-related tasks, and impediments of label-related task performance. From our findings, we derive propositions about the role of labels in OL and outline future research opportunities. Our results inform theory about the role of labels in OL and can guide practitioners leveraging labels to create and transfer knowledge within organizations.

Keywords: Knowledge Management, Organizational Learning, Machine Learning, Human-Computer Interaction.

1 Introduction

Organizations increasingly use data-driven approaches such as machine learning (ML) methods to improve organizational processes. For example, in industrial manufacturing, ML models analyze hundreds of produced parts and predict their quality (Suschnigg et al., 2020); in engineering, ML models use multi-criteria alternatives for rotor designs and suggest the most promising ones to human experts (Cibulski et al., 2020); or, in traffic control, models based on transportation data predict traffic flows (Boukerche and Wang, 2020). All these ML approaches have one essential key ingredient in common to train and evaluate sophisticated supervised and reinforcement learning models: *labeled* input data (hereafter referred to as labels) (Afiouni, 2019) and features (Bishop, 2006).

This study focuses on labels in the context of supervised and reinforcement learning. Such labels are typically created by humans and describe data categories into which data sets are classified. A simple example of a label is the tagging of animal pictures with the corresponding label "*cat*" or "*dog*". Depending on the domain of interest, labeling can depend on expert knowledge and become a complex task. An example is a diagnosis based on CT scans (Fatima and Pasha, 2017) that only specially trained experts can perform. When an expert provides a label, the knowledge about this data is retained in the label. For instance, a label related to cells can provide knowledge about the analysis subject such as "*malaria*" or "*healthy*" cells in the image (Morang'a et al., 2020). As such, labels represent an externalized knowledge product of complex expert knowledge. Thus, organizations can leverage labels not only to create ML models but to create and transfer knowledge and support organizational learning (OL).

OL is the "dynamic process of creating new knowledge and transferring it to where it is needed and used, resulting in the creation of new knowledge for later transfer and use" (Kane and Alavi 2005, p. 796). More recently research stressed the unique capabilities of ML for OL such as that learning happens within the ML model and not only through the model (Afiouni, 2019). However, for the ML model to learn, it—in complex domains—highly depends on the knowledge of domain experts. Hence, to develop ML models, data analysts need to engage with domain experts for mutual learning processes because they need to maintain relevancy to the domain and produce knowledge independently. These processes can inform a hybrid practice where ML models and domain experts together perform knowledge work. To build models, ML developers need to understand complex domains, such as manufacturing processes (Eirich et al., 2022a) by consulting experts while domain experts need to understand the decisions of an ML model (van den Broek et al., 2021). The outcome or smallest unit of this mutual learning process are labels. The relevancy of labels is undisputed since the quality of ML models depends on the quality of labels. However, ML research thus far mostly discussed labels as a means to train ML models. We want to dive deeper and understand the role of labels in OL. In particular, we aim to answer the following research question:

How do labels affect organizational learning?

In answering the research question, we conducted a case study in collaboration with a German automotive manufacturer. We draw on the theoretical framework of OL by Argote and Miron-Spektor (2011) and contextualize it to the use of labels in organizations. In particular, we analyze how labels that are used in three visual analytics (VA) tools (Thomas and Cook, 2005) affect organizational members' tasks and knowledge retention. Our results contribute a contextualized theoretical framework, which helps to understand the role of labels in OL. Furthermore, we provide propositions on how labels affect OL and outline possible future research opportunities in this context.

2 Related Work

In the following section, we explain labeling in the context of ML. This section is followed by a description of the OL framework, which serves as an analytical tool for our analyses. Subsequently, we summarize information systems literature at the intersection of ML and OL.

2.1 Labeling

In the context of ML, a label refers to attaching a certain attribute to an instance in a data set. Examples are class labels, relevance scores, similarity judgments (Bernard et al., 2018b), or the annotation of specific data spaces in observed data, such as parts of curves of polygons in images (Wang and Hua, 2011).

Labels are of categorical or numerical nature (Weber et al., 2016). In ML models, categorical labels are more common than continuous labels (Bernard et al., 2018a) and can either be scaled nominally or ordinally (Stevens, 1946). Nominally scaled labels represent discrete units of analysis, which neither have a quantitative value nor can they be ordered (e.g., male vs. female). Ordinal scaled labels, in contrast, are discrete but ordered units of analysis, where the distance between classes is vague (e.g., fast vs. slow) (Mann and Lacke, 2013). Furthermore, categorical labels can be of a binary or multi-valued nature (Bernard et al., 2018a). While binary labels allow simple user feedback, such as "*ok*" vs. "*not ok*", multi-valued labels enable broader often more specific tagging of different classes of a data instance, such as "*dog*" vs. "*cat*" vs. "*frog*". Continuous labels are based on interval or ratio scales (Stevens, 1946). Interval scales represent ordered units of analysis with the same difference (e.g., temperature). Ratio scales are equal to interval scales but contain an absolute zero value (e.g., height) (Mann and Lacke, 2013).

In the domain of ML, a label is the result of human analysis. This analysis can vary in its complexity. In very complex domains, specific domain knowledge is required to create a label such as in labeling data in the production of highly integrated electrical engines (Eirich et al., 2021) or cell proliferation (Berg et al., 2019). Less domain knowledge is required in simple analyses such as for the classification of animals (Amershi et al., 2014) or handwritten digits (Bernard et al., 2018a).

2.2 A Framework of Organizational Learning

OL focuses on the *dynamic* processes through which knowledge is created and consumed within organizations (Vera and Crossan, 2003). It can broadly be divided into two forms: exploitation and exploration (Kane and Alavi, 2005). The former refers to incremental learning, which focuses on the diffusion, reuse, and refinement of existing knowledge (Larsson et al., 1998; March, 1991; Smith and Zeithaml, 1996), whereas the latter involves the development of new or the replacement of existing knowledge in organizations (Abernathy, 1978; March, 1991; Pentland, 1995).

In line with the notion of exploration and exploitation of knowledge in organizations, Argote and Miron-Spektor (2011) provide a conceptualization of OL (see Figure 1). The key elements comprise task performance and experience, knowledge, and the active context; the latter depends on the latent organizational context. The framework provides an analytical representation of learning in organizations in a cyclic relationship between the three key elements, in which task performance and experience create or transform knowledge through interaction with the context. While the environmental context represents elements outside an organization (e.g., suppliers or customers), the latent context represents elements within the organization (e.g., a culture of trust in technologies). Although the latent context does not "*act*" actively, it affects learning through its influence on the active context. In particular, within the active context organizational members (e.g., employees) and tools (e.g., ML-based systems) perform actions, that is, tasks related to their jobs.

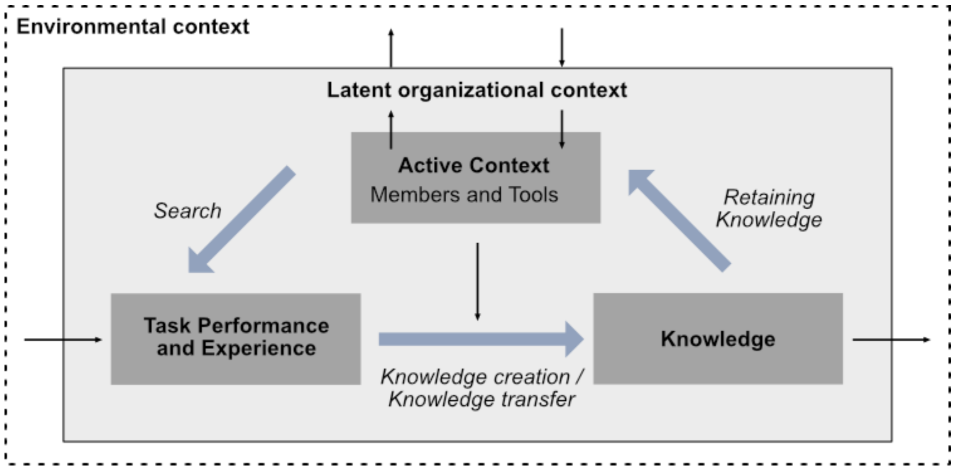


Figure 1. A framework of organizational learning based on Argote and Miron-Spektor (2011, p. 1125)

The process in which the active context affects task performance and experience is defined by Argote and Miron-Spektor as search. Knowledge is acquired through new experience (knowledge creation) or the sharing of existing experience within the organization (knowledge transfer) (Kane et al., 2005). This knowledge can either be of tacit or explicit nature (Nonaka and Takeuchi, 1995). While tacit knowledge is bound to the individual and difficult to communicate, explicit knowledge can be readily accessed by members or tools. Knowledge is retained through the active context, where organizational members or tools use it to act. Organizational members and tools, hence, represent knowledge repositories (Alavi and Leidner, 2001) and generate new experiences by performing new tasks.

2.3 ML in the Context of Organizational Learning

Research on the use of ML models in OL stresses that ML shifts the focus from the human learner to the machine. In contrast to traditional knowledge management tools (e.g., expert systems), ML technologies not only support human learning but can learn autonomously or learn in form of a hybrid practice where humans and algorithms inform each other (Sturm et al., 2021; van den Broek et al., 2021). In particular, in supervised ML models, human knowledge is crucial to inform human-in-the-loop practices, where domain experts provide inputs for algorithms, for training or debugging models, and for making sense of the results. These supervised and interactive learning approaches not only allow integrating human domain knowledge in the design of the ML model but require a huge amount of labeled data provided by human experts.

However, human-generated labels by experts can be error-prone. In knowledge-intensive domains characterized by uncertain or ambiguous knowledge work such as in generating diagnosis outputs, expert-generated labels may rely on the opinion of a single person and lack external validation. Regarding labeling tasks in such domains, the difficulty of identifying false negatives may be exacerbated because the outcome of the diagnosis may only be validated long-term, taking months or years (Lebovitz et al., 2021). Hence, especially in situations where an objective ground truth is missing, it is crucial to diligently monitor and assess the quality of labels and the ML output to avoid damage due to wrong decisions based on incorrect predictions of ML models.

In fact, a key challenge of a hybrid human-machine learning process that builds on feedback is identifying and understanding the gap between the actual and reference output. Any supervised and reinforcement ML model, hence, requires human auditing when being built, which, in turn, requires a reference measure (i.e., the ground truth provided by the domain expert) against which the output is compared. Therefore, data analysis and domain experts need to closely collaborate to alter and audit the ML model and improve its performance over time (Grønsund and Aanestad, 2020).

While research in this area focused on the development and adoption of ML models for organizational learning and its intricacies (Grønsund and Aanestad, 2020; van den Broek et al., 2021), we know little about the role of labels in this process. In response to this gap, this study investigates how labels influence OL in the context of a car manufacturer, revealing challenges in building and use of labels as well as the effects of labels (and related tools) on the key elements of OL.

3 Methodology

A case study is an empirical research approach to analyze a specific phenomenon within its environment. We chose a case study approach to study the role of labels for OL because case study designs are useful for studying context-rich sociotechnical systems (Yin, 2014), when the status quo in research is not yet well developed, when the examination of context and dynamics are important (Darke et al., 1998), and when the investigated process is difficult to observe from an outside perspective. We followed a single-case study approach (Yin, 2014). In particular, we investigated the case of a car manufacturer who introduced three ML-based VA systems building on labels. The case is situated in the context of a transformation project that many organizations currently pursue to change into a digital company.

3.1 Context and Scope

The organization in our case study is a leading German automotive manufacturer (hereafter referred to as “AutoCorp”). Like many of its competitors, AutoCorp is currently investing considerable resources in the digitization of its manufacturing processes. Therefore, dedicated analytics and engineering teams work collaboratively on the development of sophisticated VA tools to improve car production.

The two expert groups and the analyzed VA tools (see Table 1) represent our units of analysis. We refer to the two expert groups as *domain experts* and *data analysts*:

Domain Experts are responsible for specific manufacturing parts (e.g., rotor shafts of electrical engines) and have a deep understanding of a part’s physical properties (e.g., sound propagation inside electrical engines). These experts are responsible for the development of test benches, which are dedicated stations for automated part testing. Since test benches have to test a wide spectrum of a part’s properties, such as the electrical behavior of an engine as well as its mechanical attributes, domain experts also have deep knowledge about the part and its properties. For instance, two of our study participants are experts on the propagation of sound inside electrical engines. They know how an engine performs under specific conditions, for instance, in a test bench (e.g., by analyzing its magnetic field) or an electrical vehicle (e.g., by observing its acoustic signature during driving). Due to their specific expertise, domain experts are the only members of AutoCorp, who can create labels.

Data analysts develop data analysis methods, such as the training of sophisticated ML or statistical models. Their work relies on labels as they need categorical or numerical labels from domain experts as input for developing ML models or label-based tools. Thus, the more labels are available, the easier it is to develop ML models and label-based tools. Data analysts are also responsible for the acquisition and processing of data. These tasks can range from connecting sensor equipment that records measurements to machines to querying datasets from existing organizational databases.

To provide domain experts with a means to improve the manufacturing process, data analysts are currently involved in developing VA tools. Such tools aim to support the analysis of large and often unstructured datasets using visual representations and abstractions of the data (Thomas and Cook, 2005). Visualizing the data with specifically designed visualization interfaces helps experts to identify patterns and derive decisions. Thus, VA tools help in making better, data-driven decisions and—by storing labels and visualizing the data—retaining knowledge in organizational knowledge bases.

VA is closely related to the discipline of *business intelligence* (BI). Similar to VA, BI uses predefined methods (e.g., data aggregation or filtering) to gather, analyze, transform, and abstract data into new information via visualization interfaces to inform business decisions (Shollo and Galliers, 2016). Nonetheless, VA and BI differ in the following aspects: VA tools are tailormade visual interfaces between human experts and ML models to support highly explorative analyses. Examples are the analysis of speech data with self-organizing maps (Sacha et al., 2018) or root cause analyses of manufacturing errors with causality graphs (Eirich et al., 2022a). Most VA tools are designed to store data in the form of formalized user feedback, which can range from simple text inputs to more abstract forms such as graphs, to continuously develop a VA tool's underlying model. In contrast, traditional BI tools focus on the visualization of data and do not store expert feedback in dedicated databases.

Tool Name	*Description*	*Context and aim*	*Role of Labels*
RfX (Random Forest Explorer)	Visualizes the decision-making process of a random forest.	The interpretation of a random forest requires the analysis of model properties and can typically only be performed by data analysts. *RfX* allows domain experts to interactively explore the properties of a random forest and thus create new knowledge by understanding its decision-making process.	Uses labels to train a random forest classifier.
IRVINE (Interactive Labeling)	Facilitates the analysis of acoustic signatures of electrical engines.	Acoustic signatures of electrical engines contain a complex data structure, which only a few domain experts can analyze. *IRVINE* facilitates the analysis of acoustic data and helps to externalize and share domain knowledge on acoustic data in the form of labels.	Stores labels as a result of an expert's analysis.
ManEx (Manufacturing Explorer)	Allows analyzing sensor data from parts across the manufacturing process.	Manufacturing data is scattered across multiple data sources along the manufacturing process. *ManEx* helps domain experts to compare measurements from defective parts to error-free parts and thus detect the root cause of errors.	Uses labels to compare measurements from erroneous parts to error-free ones.

Table 1. Summary of analyzed label-based tools at AutoCorp

In our study, we explored the use of three tailormade, in-house VA tools (see Table 1). We hereafter refer to them as *label-based tools* because the VA tools are all interactive visual digital interfaces that rely on the input from experts to their underlying statistical models to *create, leverage,* and/or *store* knowledge in the form of labels. Hence, these VA tools do not only depend on domain knowledge but can also formalize that knowledge with labels through their visual interface.

3.2 Data Collection and Coding

We conducted the case study over sixteen months. During this period, the first author was deeply involved in the development of all label-based tools. All study participants are current users of the label-based tools. To develop and evaluate each tool, we collected data from the following sources:

(1) *Semi-structured interviews*: After the roll-out of each label-based tool, we interviewed 15 participants—seven data analysts (DA-1 to DA-7) and eight domain experts (DE-8 to DE-15) of AutoCorp. All interviewees were involved in the development of the tools and are current users of each label-based tool. They are all male, were on average 34.2 years old, had a mean working experience in AutoCorps of 9.7 years, and did not use label-based tools before the study. Each interview took on average 33 minutes.

The interview guideline was developed based on our theoretical lens, i.e., the effect of labels on OL. First, considering the dimension active context from the original OL framework (Figure 1), we asked questions on how participants define labels, the roles of the employees, and novel tools such as RfX (Eirich et al., 2022b), IRVINE (Eirich et al., 2021) or ManEx (Eirich et al., 2022a) in using labels or making decisions based on labels. Second, in the dimension task performance and experience, we asked, for instance, about the tasks that are supported by labels or which impediments exist in labeling tasks for organizational members and tools. Third, in the knowledge dimension, questions pertained to the storage of labels.

(2) *Direct observations:* The first author investigated how organizational members use the three label-based tools (RfX, IRVINE, ManEx) following Someren et al.'s (1994) Think Aloud Method. The researcher asked domain experts and data analysts to express their thoughts verbally when using label-based tools (Someren et al., 1994). The observations helped us to understand how organizational members (e.g., domain experts from different organizational units) share and create knowledge in the form of labels. Furthermore, the observations helped to contextualize label-based tasks.

(3) *Documentation from design and development iterations*: Each label-based tool, introduced in Section 3.1, was designed and developed in several iterations. During each iteration, the on-site researcher documented numerous types of information. This information included the tasks users perform while using the tools or VA requirements used as design targets for each tool. This data helped to abstract high-level system requirements, which *label-based tools* need to address. The data also informed the definition of general tasks, which members and tools perform when using label-based systems.

Following our theoretical framework, we *coded* the data in the dimensions of the OL process. In particular, we went through a constant comparison analysis (Boeije, 2002). In so doing, we grouped the data into themes and gave names to each theme to describe their content. Next, we categorized these themes into the dimensions of the OL framework. We went through this process three times to refine the coding of the data.

4 Case Study Results

The focus of this analysis is on how labels affect each dimension of the OL framework and how they circulate along its three dimensions (active context, task performance and experience, and knowledge). We summarized our findings in Figure 2, which shows our adapted version of the original OL framework (Argote and Miron-Spektor, 2011).

Additionally, we asked experts how they define labels. From the answers, we derived the following definition:

> *A label is a categorical or numerical result of a complex data analysis performed by domain experts to describe a product's quality. It is, hence, the representation and aggregation of expert knowledge.*

Although this definition might not apply to contexts, in which expert knowledge is unnecessary, it summarizes the views of AutoCorp's experts and thus, will be used in the subsequent sections. In the active context (Section 4.1), labels affect both *members* and *tools* to perform the tasks *provide* and *decide* (Section 4.2). When performing these tasks new *tacit* and *explicit* knowledge (Section 4.3) is created.

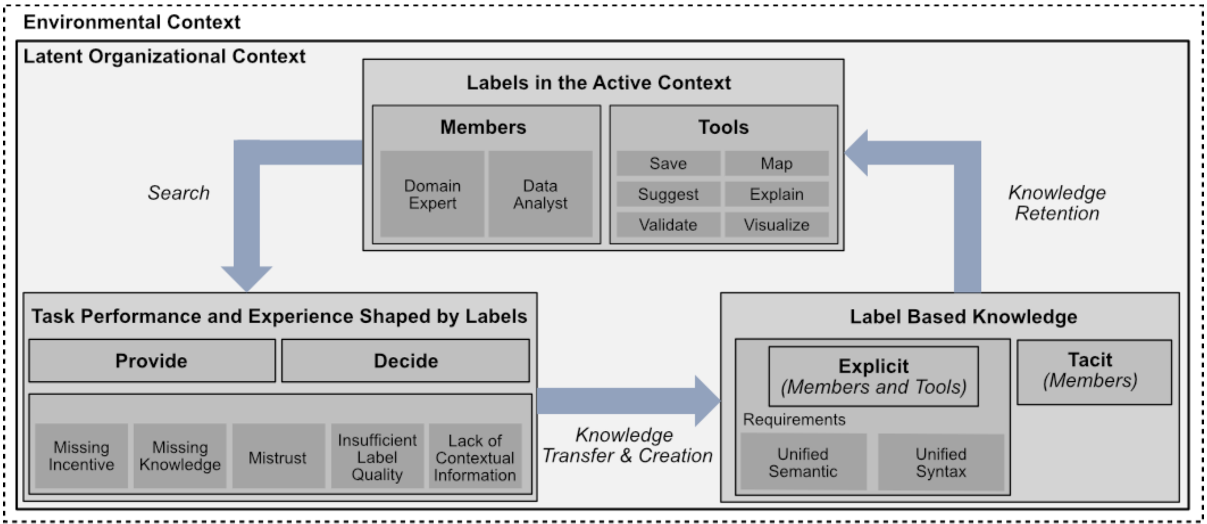


Figure 2. The organizational learning framework contextualized to labels

4.1 Active Context

As outlined in Section 3.1, members who work with labels are either *domain experts* or *data analysts*. Furthermore, the active context of the framework includes *label-based tools*, which are developed by data analysts in close collaboration with domain experts.

Labels in the active context represent a semantic, which facilitates the collaboration between domain experts and data analysts. As such, labels capture and hence, serve to share explicit knowledge. Specifically, domain experts provide and share labels with data analysts. Data analysts can use these labels to develop label-based tools or to better comprehend the problems that domain experts are facing. Label-based tools also allow domain experts to increase their analytical reasoning and understanding because they provide an interactive means to compare complex data structures, such as the production quality of different manufacturing stations. Thus, label-based tools allow members to build on each other's knowledge and facilitate OL. Hence, we propose:

P1a: *Labels representing explicit knowledge provide a semantic for organizational members which positively affects the comprehension of knowledge-intensive domains and thus positively impacts OL.*

Based on the experts' opinion of the three label-based tools, we derived the six high-level requirements *save, map, suggest, explain, validate*, and *visualize* that these tools should address.

Save: Label-based tools should store labels in organizational databases. This is necessary to make labels readily accessible for analyses by other organizational members within the same or different domains. IRVINE, for instance, allows saving labels from analyses of acoustic data from electrical engines. These labels are also relevant for domain experts outside the domain of acoustics. Engineers, who are responsible for the development of novel generations of electrical engines, can use stored labels to reflect on the design of electrical engines and include this externalized knowledge in future part generations.

Map: Label-based tools should provide information about how labels are stored in different organizational databases. For example, at each step of the manufacturing process, measurements for engines are recorded and stored either locally or scattered across organizational cloud infrastructures. Domain experts reported that it is "*extremely important to harmonize measurements and labels from different steps to understand the relation between them across the manufacturing process*" (DE-8). In this regard, DE-10 explained why he uses the tool ManEx: "*Fully comprehending the relation of measurements and groups of labeled engines is crucial to find the root cause of an error as early as possible in the manufacturing process.*"

Suggest: Tools should make recommendations about analysis objects (e.g., electrical engines) that are most suitable for labeling. Especially when members analyze large numbers of data, this requirement is necessary to reduce the number of analysis objects from the plethora of available options. IRVINE addresses this requirement by suggesting groups of most relevant engines for an in-depth analysis using ML predictions reducing the time domain experts need to find and analyze particularly relevant engines by over 30%. ManEx addresses this requirement by guiding domain experts and data analysts to groups of engines with the most deviating measurements along the manufacturing process, which helps members to easily locate and label engines with previously unnoticed errors.

Explain: Since label-based tools often address high-cognition tasks, they also should provide additional information about their decision-making process (Rudin, 2019). This helps organizational members to better understand a tool's decision, for instance, a prediction about the quality of a part. DA-2, for example, argued that label-based tools "*which use ML models should provide data so I can understand the reason why it predicted a specific error for a particular engine."* The explanation of a model depends on the label type. For example, RfX visualizes predictions of multiple decision trees in a random forest for categorical labels as a node-link diagram. In addition, IRVINE outlines regions of interest in time series data of acoustic signatures.

Validate: Label-based tools should provide a quality measure for labels. This is important to ensure consistent and standardized labeling processes, which, in turn, allow tools to continually learn from labels. ManEx, for instance, addresses this requirement by providing a validation mechanic where domain experts can evaluate how different measurements influence groups of labeled engines. Furthermore, RfX shows different quality metrics for ML models results, which help to evaluate whether more labels are needed to train models or whether the ML model achieves a sufficient prediction accuracy. DE-14 argued to establish even more validation mechanisms, for example, that "*labels must be validated by multiple experts*" or that "*tools should include wrong predictions as an additional safety check, which need to be detected by an expert.*"

Visualize: Label-based tools should include an interactive visualization interface (Thomas and Cook, 2005). A visualization interface is necessary to relate labels to associated data by giving them a visual context. This visualization should enable data analysts and domain experts to comprehend the data or to analyze patterns, trends, or errors in large datasets easily. Although all three tools introduced in the case description include visualization interfaces, this requirement is undisputed for all new tools, which use labels. DE-8 explains why visualization is so important: "*If I cannot see the model predictions embedded in a visualization interface, I find it hard to make a decision based on the prediction.*"

Search: In the original OL framework (see Figure 1), *search* is the sub-process that connects the active context with task performance and experience. To perform a task, members and tools need to effectively locate labels and "*need to know where to find them*" (DA-3). The easier the locatibility of labels is, the easier it is for members and tools to perform a label-related task. Data analysts need to know which type of label exists in which database and how to write queries for each database to retrieve labels. Domain experts need to know the location of specific file storages to search for labels more efficiently.

Label-based tools, which address the high-level requirements, can increase the knowledge of domain experts and data analysts by supporting locating and analyzing interesting data instances. For instance, label-based tools that suggest new data instances for labeling reduce the overall amount of data domain experts must analyze to a subset of particularly interesting data. These subsets may represent engine parts with highly anomalous sensor signals. Such signals, in turn, are particularly helpful in detecting and understanding new types of possible errors in the manufacturing process and thus foster the creation of new knowledge. Hence, we propose:

P1b: *Label-based tools that save, map, suggest, explain, validate, or visualize labels positively affect OL through their property to easily locate and analyze data instances of interest for human analyses.*

4.2 Task Performance and Experience

In the dimension *task performance and experience*, we surfaced the two label-related tasks *provide* and *decide* that label-based tools and members frequently perform.

In the task **provide**, domain experts create labels by a manual analysis that can be supported by label-based tools, such as IRVINE. An example of a manual labeling task is the analysis of time-series measurements, in which a label is created by assigning a property to the measurement. For example, DE-10 explained how he manually provides labels: "*When the time-series surpasses a certain threshold, I assign an error to the data*". Furthermore, providing labels is supported with designated labeling systems such as IRVINE. IRVINE facilitates the labeling of multiple engines by clustering engines with similar data patterns and presenting them to users for an in-depth analysis.

In the task **decide**, members and tools process labels for a specific analysis of manufacturing data and make a decision based upon this analysis. Domain experts perform analyses with labels, for instance, "*to find the exact location of an error in a labeled part*" (DE-11) or "*to understand the cause of an error*" (DE-8). Labels support decision-making tasks by "*providing a direction that points towards a decision*" (DA-6). Domain experts, for instance, need to decide on "*the adaptation of an existing testing procedure*" (DE-12). In contrast, data analysts, for example, decide based on "*a model recommendation, which electric vehicle is best suited for an in-depth analysis*" (DA-7). The tool ManEx automatically decides when to notify domain experts. For instance, in case of an unexpectedly high number of errors in the manufacturing process, the model will alert an expert about the errors.

Labels shape task performance and experience when the respective tasks either rely on labels or are strongly influenced by labels. For example, labels can be used to improve label-based tools to make better suggestions of new relevant data instances for labeling. During the process of providing labels, domain experts gain a better understanding of the data they analyze and thus, create new knowledge, which can be externalized in the form of labels. In terms of the task "decide", labels can improve decision-making and thus result in new knowledge. For example, label-based tools can proactively inform domain experts when produced parts contain a large number of anomalies in their sensor data. Domain experts then need to decide whether to send these parts to reworking stations or not. By taking this label-based decision, domain experts gain a better understanding of sensor data in the manufacturing process resulting in new tacit knowledge. Hence, we propose:

P2a: *Label-based tools support domain experts in providing labels and thus have a positive impact on the storing and sharing of explicit knowledge in organizations.*

P2b: *Label-based tools positively affect decision-making capabilities and thus the creation of new knowledge.*

In addition, we identified the five impediments *missing incentives*, *missing knowledge*, *mistrust*, *insufficient labels quality*, and *lack of contextual information*, which can hinder task performance:

Missing incentives is a lack of extrinsic or intrinsic motivation of a domain expert to provide a label. DE-13, for instance, pointed out that he does "*not have enough time to manually analyze hundreds of engines and provide a label for every single one.*" Missing incentives, however, negatively affect the work of data analysts, who rely on labels. To tackle this impediment, data analysts try to run analyses with few labels only. However, these analyses do not provide good results as DA-4 explains: "*With few labels, I can only run simple statistical models, which give little insights about the cause of an error.*"

Missing knowledge is a lack of understanding the domain (e.g., the exact functionality of a manufacturing station) by a data analyst or domain expert to perform the tasks provide and decide. By missing knowledge, we mean a lack of knowledge about specific domain problems and not the use of label-based tools. DA-4 explained how missing domain knowledge slows down his work: "*I first have to fully comprehend the problem domain and understand how to create labels, which is a very tedious process.*" In addition, DE-13 found it "*particularly difficult to create new labels or use existing labels if the problem to solve is new and I do not have the necessary knowledge to understand it.*"

Mistrust is a suspicious attitude of data analysts or domain experts when collaborating with other members or working with label-based tools (Thiebes et al., 2021). Trust often relies on personal relationships between members who create and use labels. DE-12 explained how trust in relationships affects his work with labels: "*I have more trust in a label if it is created by a person I know than in a label from a person I do not know. If I do not know the person, I always crosscheck the label myself.*" Trust also relies on a member's perception of a domain expert's or data analyst's expertise. A data analyst's ability to explain the statistical results of an analysis to a domain expert is an example of that. In this context, DA-5 explained: "*If I cannot respond to the question of why I believe that the specific anomaly is significant, the person for whom I have to perform the analysis does not trust my analysis.*" Also, a lack of trust in *label-based tools* can harm task performance as DE-13 noted: "*I need some kind of information on the trustworthiness of the model's prediction, for example, a confidence interval. Lacking this information, I find it difficult to make an informed decision based on a prediction.*"

Insufficient label quality refers to a label that captures false information about a part. Examples are *false positives*, such as an error-free part labeled as erroneous, or *false negatives,* such as an erroneous part labeled as error-free. Analyses based on false labels can result in incorrect analyses results of data analysts and domain experts or false predictions by label-based tools. Decisions based on wrong results can have severe negative effects on the work of members and the quality of parts. DA-4 provided the following example: "*Someone presented an ML model and all thought that it performed well. However, after a while, the model did make wrong predictions. It took us a lot of time to identify the reason, which was that it was trained with low-quality labels.*"

Lack of contextual information refers to the missing but necessary information to interpret a label correctly. Examples of such information are the "*location where the label was produced*" (DA-1), "*who provided the label*" (DE-12), "*the timestamp of a label*" (DE-10), or missing/incomplete data (Gashi et al., 2021) as described by DA-7 as a *"lack of relevant sensor data for a label.*" If labels do not include relevant context information, members have difficulties in fully comprehending the meaning of a label as DE-15 stated: "*I need to have additional information, for example, on which day the label was provided, to know what it means.*"

Knowledge creation and transfer connects task performance and knowledge. While performing the two label-related tasks (provide and decide), members gain new experiences and learn to perform tasks more efficiently. For instance, by discussing individual decision trees visualized by RfX, we observed that members started to develop a shared meaning for a random forest. A shared meaning of this ML model, in turn, enhances the collaboration between domain experts and data analysts to improve RfX. P1 provides another example of how he creates new knowledge with labels: "*The more I work with labels, the better I comprehend the manufacturing process and gain new knowledge about it.*" Labels also help to transfer knowledge between members and tools; DE-15 stated: "*Labels help to exchange information since I can pass labels to my colleagues so they can use them for their analyses*."

The aforementioned impediments can hinder the successful execution of the tasks *provide* and *decide*, and hence the creation and transfer of knowledge. For example, missing incentives or knowledge to provide labels result in fewer labels provided by domain experts. Hence, less explicit knowledge captured in labels circulates inside the organization. Furthermore, mistrust, insufficient label quality, and a lack of contextual information may result in the negative attitude of members towards label-based tools (Chatzimparmpas et al., 2020) and thus, negatively affect knowledge transfer and creation through labels. Hence, we propose:

P2c: *Missing incentives and knowledge to provide labels, mistrust in tools, insufficient label quality, and a lack of contextual information negatively affect knowledge transfer and creation.*

4.3 Label-based Knowledge

For the category knowledge, we surfaced the two knowledge types, *tacit* and *explicit knowledge*, in the context of creating and using labels:

Tacit knowledge is created in the interaction between members or between members and label-based tools. DE-8, for instance, creates knowledge by sharing experiences or discussing labels with other experts: "*Labels help me in discussions, where my colleagues and I compare error-free parts to erroneous parts and collaboratively find the causes of an error.*" Members can also interpret the representation of labels and associated data, for example, by using the tools RfX, IRVINE, or ManEx and thus continuously learn from them.

Explicit knowledge is captured in labels (Bernard et al., 2018a), which can be stored and accessed by organizational members or tools. For instance, when a domain expert creates a label with IRVINE it is saved in an organizational database. Labels can then be used to develop ML models as DE-9 noted: "*When I run the same analyses with new labels, I can evaluate the quality of my analysis process over time and gain new interesting insights.*" Furthermore, domain experts, data analysts, or label-based tools can combine labels with different data sources into more systematic and comprehensive sets of explicit knowledge (Wang et al., 2009).

Both, a *unified semantic* and *unified syntax* are important requirements for storing labels and the storage of explicit knowledge.

Unified semantic refers to the description of labels, as well as their properties and relations by defining a set of abstract concepts and categories that represent and relate labels. This is necessary so that labels are interpreted the same way by different members of AutoCorp. DE-15 provided one approach to build a unified semantic: "*different domain experts should decide on a common metric, which fits best to create labels.*" Furthermore, DA-7 pointed out that "*we need to define a common ontology, to understand how labels are embedded in the manufacturing process.*" A unified semantic is also important for the interpretation of labels in label-based tools because "*different tools, which all process labels with a common data processing routine would make it easier to compare them*" (DE-15).

Compatible syntax refers to a common structure and data format for storing and sharing labels within organizations. In particular, labels should be available in "*numerical interchangeable data formats*" (DE-9) and be "*readable by machines*" (DE-7). A unified syntax also includes that labels are "*compatible with different kinds of data sources*" (DE-7) allowing the combination of labels with multiple datasets.

Knowledge retention connects knowledge with the active context. To continually retain knowledge in the form of labels and to transfer it from the short-term to long-term organizational memory, labels should be "*easily accessible to all individuals, who need to work with the labels*" (DA-1). Hence, labels need to be stored in a central database. Nevertheless, labels that contain sensitive organizational knowledge "*should only be available to specific members who have to work with these labels*" (DE-12). Sensitive knowledge can pertain to labels about prototypical parts, which the organization wants to keep secret. In the case of sensitive data, label databases should contain only restricted user access.

That being said, for the efficient retention of label-based knowledge in an organization, labels must be stored uniformly with a unified semantic and syntax. IS researchers have stressed that the standardization of knowledge is crucial to improve organizational knowledge retention (Förderer et al., 2014; Hsiao, 2008). To implement a unified semantic, standardization methods for labels such as ontologies (Alvarez-Coello and Gomez, 2021) or knowledge graphs (Ehrlinger and Wöß, 2016) can be used. To make labels readily accessible across organizational units, a common syntax, for example, compatible data types of labels are necessary. Hence, we propose:

P3a: *A unified semantic for labels positively affects the retention of tacit and explicit knowledge.*

P3b: *A compatible syntax for labels positively affects the retention of tacit and explicit knowledge.*

5 Discussion

To answer our research question and to gain a better understanding of how labels influence the process of OL, we adapted Argote and Miron-Spektor's (2011) theoretical OL framework (see Figure 2). Drawing on our results, we propose a set of research questions summarized in Table 2.

Since labels present explicit knowledge (Wang et al., 2009), they play an important role in OL and complement well-established explicit but more generic knowledge products such as images or symbols (Peltokorpi et al., 2007). In contrast to such generic knowledge products, labels and related tools explicitly shape domain experts' routine work processes and help to better understand certain domains. Hence, sharing labels in organizations can have a positive effect on organizational learning since members can use them to understand knowledge-intensive domains. In our study, we analyzed the role of labels inside a single organizational unit. Since the original process model of OL (Kane et al., 2005) is also about OL across organizational units and between organizations, our first RQ proposes to understand how labels can improve knowledge creation and transfer in these broader contexts (RQ1).

Propositions	*Future research opportunities (RQ)*
1a and 1b	*RQ1*: How can labels efficiently be used across different organizational units and between organizations? *RQ2*: How must label-based tools be designed to improve knowledge creation and transfer across different organizational units and between organizations?
2a, 2b, and 2c	*RQ3*: How can organizations design incentives to foster label-based tasks and roles? *RQ4:* How can organizations overcome impediments related to labels that hinder effective task performance? *RQ5*: Is a data labeler a new position that organizations need to incorporate? And if so, what are the tasks and what knowledge is required of a labeler?
3a and 3b	*RQ6*: How must labels be modeled in organizational ontologies to support organizational knowledge management? *RQ7*: How can ontologies support the development of label-based tools?

Table 2. Future research opportunities

A further relevant future research direction is the design and development of label-based tools to support the efficient sharing of labels inside and between organizations (RQ2). In so doing, we believe that the two stakeholder groups—domain experts and data analysts—are crucial. For instance, data analysts should not only consider what data to include when building label-based tools but also how domain experts can provide labels during the development of label-based tools. Furthermore, our high-level requirements of label-based tools can guide researchers and practitioners with the design of similar decision support systems, such as the presented VA tools or similar systems, such as BI.

In addition, we surfaced impediments, which hinder the execution of the tasks provide and decide. One impediment refers to missing incentives to provide labels. Hence, we suggest investigating how organizations could design incentives to foster label-based tasks and roles to overcome such impediments (*RQ3*). One strategy to increase the engagement of domain experts in these tasks could be to apply gamification (Khakpour and Colomo-Palacios, 2020) or rewards for labeling activities similar to rewarding inventions or patents (Giarratana et al., 2018).

In line with prior research in complex and uncertain domains where high uncertainty and lack of validity of labels can lead to wrong predictions in ML models (Lebovith et al. 2021), we found that a lack of domain knowledge can affect the quality of ML models. Organizations need to develop strategies to overcome such issues, for instance, by flagging labels that are uncertain as well as continuously monitoring, assessing, and developing the ML model over time *(RQ4)*.

Concerning the task *provide*, a new organizational role similar to established organizational roles, which deals with data-related tasks (Crisan et al., 2020) could be introduced. Crisan et al. (2020) summarized existing roles, such as the "*data engineer*", who is mainly responsible for the development of data gathering and processing tasks. In addition to the roles defined by Crisan et al. (2020), organizations may consider introducing the "*data labeler*" (*RQ5*). This role could include organizational members, who have not only knowledge about a specific domain such as electrical engineering but are also responsible for providing high-quality labels for label-based tools as part of their job descriptions.

One way to represent knowledge in organizations should be in a machine-interpretable form. In the field of artificial intelligence, researchers investigated the role of knowledge representation with ontologies (Studer, 2007) to provide ML models with conceptual abstractions for particular domains of interest (Gonçalves et al., 2019; Harispe et al., 2014; Smaili et al., 2019). Future research may focus on the role of labels as knowledge products in the design of ontologies for organizational knowledge representation (RQ6). When a unified semantic and compatible syntax for labels in organizational knowledge bases are established, another future research avenue is to understand how these affect the creation of label-based tools (RQ7). In particular, we assume that access to labels structured by a compatible syntax and unified semantic will foster the development of label-based tools by reducing the need to cross-check labels with domain experts or to convert and harmonize labels into a consistent data format.

6 Limitations

We acknowledge that our study has limitations. First, the study is based on qualitative data drawn from analyzing the use of three label-based tools, participant observations, and interviews with experts who work all in the same automotive company. Hence, our results are limited to the context of our study and the results may have limited validity in other domains. Nevertheless, since labels are such a basic data ingredient for any ML model, also other domains may reveal similar patterns regarding the circulation of labels between the three dimensions of the OL framework.

Second, our adapted OL framework does not address the latent organizational context nor the environmental context of the original framework by Argote and Miron-Spektor (2011). This was because in AutoCorp the role of labels is relatively new and label-based tools are not yet institutionalized. Further research can build on our propositions and research opportunities to investigate how labels affect OL.

7 Conclusion

In this paper, we analyze the impact of labels on OL. In so doing, we conducted a case study in collaboration with a German car manufacturer to understand how labels can support OL. The result is an adapted framework of OL focusing on the role of labels in OL that informs about organizational members who work with labels, requirements of label-based tools, label-related tasks and impediments of task performance, and how all these affect OL within an organization. Furthermore, we outline seven propositions about the role of labels regarding OL and suggest possible label-related future research directions. Practitioners can use the framework to understand labels along different stages of OL and to improve learning and knowledge management within their organizations.

References

Abernathy, W.J., 1978. The productivity dilemma: roadblock to innovation in the automobile industry. Johns Hopkins University Press, Baltimore.

Afiouni, R., 2019. Organizational Learning in the Rise of Machine Learning, in: International Conference on Information Systems.

Alavi, M., Leidner, D.E., 2001. Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly 25, 107-136.

Alvarez-Coello, D., Gomez, J.M., 2021. Ontology-Based Integration of Vehicle-Related Data, in: IEEE International Conference on Semantic Computing.

Amershi, S., Cakmak, M., Knox, W.B., Kulesza, T., 2014. Power to the People: The Role of Humans in Interactive Machine Learning. AIMag 35, 105–120.

Argote, L., Miron-Spektor, E., 2011. Organizational Learning: From Experience to Knowledge. Organization Science 22, 1123–1137.

Berg, S., Kutra, D., Kroeger, T., Straehle, C., Kausler, B., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Köthe, U., Hamprecht, F., Kreshuk, A., 2019. ilastik: interactive machine learning for (bio)image analysis. Nature Methods 16, 1–7.

Bernard, J., Hutter, M., Zeppelzauer, M., Fellner, D., Sedlmair, M., 2018a. Comparing Visual-Interactive Labeling with Active Learning: An Experimental Study. IEEE Transactions on Visualization and Computer Graphics, 298-308.

Bernard, J., Zeppelzauer, M., Lehmann, M., Müller, M., Sedlmair, M., 2018b. Towards User-Centered Active Learning Algorithms. Computer Graphics Forum 37, 121–132.

Bishop, C.M., 2006. Pattern recognition and machine learning, Information science and statistics. Springer, New York.

Boeije, H., 2002. A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews. Quality and Quantity 36, 391–409.

Boukerche, A., Wang, J., 2020. Machine Learning-based traffic prediction models for Intelligent Transportation Systems. Computer Networks 181, 107530.

Chatzimparmpas, A., Martins, R.M., Jusufi, I., Kucher, K., Rossi, F., Kerren, A., 2020. The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations. Computer Graphics Forum 39, 713–756.

Cibulski, L., Mitterhofer, H., May, T., Kohlhammer, J., 2020. PAVED: Pareto Front Visualization for Engineering Design. Computer Graphics Forum 39, 405–416.

Crisan, A., Fiore-Gartland, B., Tory, M., 2020. Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers, in: IEEE Visualization in Data Science.

Darke, P., Shanks, G., Broadbent, M., 1998. Successfully completing case study research: combining rigour, relevance and pragmatism. Information Systems Journal 8, 273–289.

Ehrlinger, L., Wöß, W., 2016. Towards a Definition of Knowledge Graphs, in: SEMANTiCS.

Eirich, J., Bonart, J., Jäckle, D., Sedlmair, M., Schmid, U., Fischbach, K., Schreck, T., Bernard, J., 2021. IRVINE: A Design Study on Analyzing Correlation Patterns of Electrical Engines. IEEE Transactions on Visualization and Computer Graphics, 11-21.

Eirich, J., Koutroulis, G., Mutlu, B., Jackle, D., Kern, R., Schreck, T., Bernard, J., 2022a. ManEx: The Visual Analysis of Measurements for the Assessment of Errors in Electrical Engines. IEEE Comput. Grap. Appl. 1–18.

Eirich, J., Münch, M., Jäckle, D., Sedlmair, M., Bonart, J., Schreck, T., 2022. RfX: A Design Study for the Interactive Exploration of a Random Forest to Enhance Testing Procedures for Electrical Engines. Computer Graphics Forum 1–18.

Fatima, M., Pasha, M., 2017. Survey of Machine Learning Algorithms for Disease Diagnostic. JILSA 9, 1–16.

Förderer, J., Schütz, S., Kude, T., 2014. Add-on Solution Success: A Configurational View on Knowledge Sharing in Digital Platforms, in: International Conference on Information Systems.

Gashi, M., Ofner, P., Ennsbrunner, H., Thalmann, S., 2021. Dealing with missing usage data in defect prediction: A case study of a welding supplier. Computers in Industry 132, 103505.

Giarratana, M.S., Mariani, M., Weller, I., 2018. Rewards for Patents and Inventor Behaviors in Industrial Research and Development. AMJ 61, 264–292.

Gonçalves, R.S., Kamdar, M.R., Musen, M.A., 2019. Aligning Biomedical Metadata with Ontologies Using Clustering and Embeddings, in: Hitzler, P., Fernández, M., Janowicz, K., Zaveri, A., Gray, A.J.G., Lopez, V., Haller, A., Hammar, K. (Eds.), The Semantic Web. Springer International Publishing, Cham, pp. 146–161.

Grønsund, T., Aanestad, M., 2020. Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems 29, 101614.

Harispe, S., Ranwez, S., Janaqi, S., Montmain, J., 2014. The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinformatics 30, 740–742.

Hsiao, R., 2008. Knowledge Sharing in a Global Professional Service Firm. MIS Q. Executive 7.

Kane, A.A., Argote, L., Levine, J.M., 2005. Knowledge transfer between groups via personnel rotation: Effects of social identity and knowledge quality. Organizational Behavior and Human Decision Processes 96, 56–71.

Kane, G., Alavi, M., 2005. Information Technology and Organizational Learning: An Investigation of Exploration and Exploitation Processes, in: Organization Science.

Khakpour, A., Colomo-Palacios, R., 2020. Convergence of Gamification and Machine Learning: A Systematic Literature Review. Technology, Knowledge and Learning.

Larsson, R., Bengtsson, L., Henriksson, K., Sparks, J., 1998. The Interorganizational Learning Dilemma: Collective Knowledge Development in Strategic Alliances. Organization Science 9, 285–305.

Lebovitz, S., Levina, N., Lifshitz-Assa, H., 2021. Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. MIS Quarterly 45, 1501–1526.

Mann, P.S., Lacke, C.J., 2013. Introductory statistics, 8. ed., international student version. ed. Wiley, Hoboken, NJ.

March, J.G., 1991. Exploration and Exploitation in Organizational Learning. Organization Science 2, 71–87.

Morang'a, C.M., Amenga–Etego, L., Bah, S.Y., Appiah, V., Amuzu, D.S.Y., Amoako, N., Abugri, J., Oduro, A.R., Cunnington, A.J., Awandare, G.A., Otto, T.D., 2020. Machine learning approaches classify clinical malaria outcomes based on haematological parameters. BMC Med 18, 375.

Nonaka, I., Takeuchi, H., 1995. The knowledge-creating company: how Japanese companies create the dynamics of innovation. Oxford University Press, New York.

Peltokorpi, V., Nonaka, I., Kodama, M., 2007. NTT DoCoMo's Launch of I-Mode in the Japanese Mobile Phone Market: A Knowledge Creation Perspective. Journal of Management Studies 44, 50–72.

Pentland, B.T., 1995. Information systems and organizational learning: The social epistemology of organizational knowledge systems. Accounting, Management and Information Technologies 5, 1–21.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215.

Sacha, D., Kraus, M., Bernard, J., Behrisch, M., Schreck, T., Asano, Y., Keim, D.A., 2018. SOMFlow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance. IEEE Transactions on Visualization and Computer Graphics 24, 120–130.

Shollo, A., Galliers, R.D., 2016. Towards an understanding of the role of business intelligence systems in organisational knowing: The role of BI systems in organisational knowing. Information Systems Journal 26, 339–367.

Smaili, F.Z., Gao, X., Hoehndorf, R., 2019. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. Bioinformatics 35, 2133–2140.

Smith, A., Zeithaml, C., 1996. Garbage Cans and Advancing Hypercompetition: The Creation and Exploitation of New Capabilities and Strategic Flexibility in Two Regional Bell Operating Companies. Organization Science 7, 388–399.

Someren, M.W. van, Barnard, Y.F., Sandberg, J.A.C., 1994. The think aloud method: a practical guide to modelling cognitive processes. Academic Press, London.

Stevens, S.S., 1946. On the Theory of Scales of Measurement. Science 103, 677–680.

Studer, R., 2007. Semantic Web Services: concepts, technologies, and applications, Semantic Web Services. Springer, Berlin.

Sturm, T., Gerlach, J., Pumplun, L., Mesbah, N., Peters, F., Tauchert, C., Nan, N., Buxmann, P., 2021. Coordinating Human and Machine Learning for Effective Organizational Learning. MIS Quarterly 45, 1581–1602.

Suschnigg, J., Mutlu, B., Fuchs, A., Sabol, V., Thalmann, S., Schreck, T., 2020. Exploration of Anomalies in Cyclic Multivariate Industrial Time Series Data for Condition Monitoring, in: International Workshop on Big Data Visual Exploration and Analytics.

Thiebes, S., Lins, S., Sunyaev, A., 2021. Trustworthy artificial intelligence. Electron Markets 31, 447–464.

Thomas, J.J., Cook, K.A. (Eds.), 2005. Illuminating the path: the research and development agenda for visual analytics. IEEE Computer Soc, Los Alamitos, Calif.

van den Broek, E., Sergeeva, A., Huysman Vrije, M., 2021. When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. MISQ 45, 1557–1580.

Vera, D., Crossan, M., 2003. Organizational learning and knowledge management: toward an integrative framework. Handbook of Organizational Learning and Knowledge Management 122–141.

Wang, M., Hua, X.-S., 2011. Active Learning in Multimedia Annotation and Retrieval: A Survey. ACM Transactions on Intellilgent Systems and Technologies 2.

Wang, X., Jeong, D.H., Dou, W., Lee, S.-W., Ribarsky, W., Chang, R., 2009. Knowledge Assisted Visualization: Defining and Applying Knowledge Conversion Processes to a Visual Analytics System. Computer Graphics 33, 616–623.

Weber, N., Waechter, M., Amend, S.C., Guthe, S., Goesele, M., 2016. Rapid, Detail-Preserving Image Downscaling. SIGGRAPH Asia 6.

Yin, R.K., 2014. Case study research: design and methods, Fifth edition. ed. SAGE, Los Angeles.

6.8 P8: Visual Analytics for IoT Data from Large Scale Manufacturing Processes

Full reference: *J. Eirich. Visual analytics for iot data from large-scale manufacturing processes. American Conference on Information Systems, 08 2022*

Visual Analytics for IoT Data from Large-Scale Manufacturing Processes

Completed Research Full Papers

Joscha Eirich
University of Bamberg
joscha.eirich@uni-bamberg.de

Abstract

Advances in technologies, such as the Internet of Things (IoT) and Visual Analytics (VA), are enabling a new generation of smart manufacturing. These technologies enable the efficient tracking of the quality of produced parts along every step in the manufacturing process. However, the challenge remains that distinct IoT data sources must be connected, harmonized, and made readily accessible to human experts for analyses. In this paper, we followed a design science research approach to develop a VA artifact supporting engineering experts in analyzing IoT data from interconnected stations of a manufacturing process for electrical engines. We developed our artifact in collaboration with an industrial partner from the automotive sector and evaluated it with five engineering experts. Results indicate high usability and usefulness of the artifact as part of a real-world manufacturing process. Our instantiated artifact can serve as guidance to researchers and practitioners, who work in similar manufacturing domains.

Keywords

Keywords: Visual Analytics, IoT data, Design Science, Human-Computer Interaction.

Introduction

To improve production performance and gain competitive advantages in global markets, the manufacturing industry invests in new technologies, such as the *Internet of Things* (IoT) (Yang et al. 2019). Nowadays, manufacturing processes increasingly include sophisticated sensor equipment, which results in interconnected IoT networks and hence a new generation of smart manufactories. One way to create value from IoT data and support domain experts is to use data visualization approaches (Thomas and Cook 2005). *Visual Analytics* (VA) can help to integrate human experts into the visual exploration of IoT data by supporting the analytical reasoning capabilities of experts with model-supported and custom-designed visualization interfaces (Keim et al. 2008).

Thus far, research has analyzed how domain experts can be supported in the analysis of IoT data in the manufacturing sector with VA (Post et al. 2017; Suschnigg et al. 2020). However, previous researchers only addressed isolated parts of manufacturing processes, such as single test benches (Suschnigg et al. 2020). Nevertheless, to efficiently track the quality of parts, it is necessary to study manufacturing processes as a whole system. During manufacturing processes, parts are produced at manufacturing steps (hereafter: steps) inheriting manufacturing stations (hereafter: stations), which all record IoT data. The challenge is to connect data from different sensors to a single cohesive dataset. Furthermore, measurements need to be harmonized for experts to compare them along with the different steps and stations. From a visualization perspective to the best of our knowledge, no research has been carried out addressing the problem of how to design and implement a VA system, capable of visualizing data from an entire manufacturing process. We aim to address this gap by answering the following research question:

"How must a visual analytics artifact be designed to support human experts with the analysis of large-scale IoT data from an entire manufacturing process?"

To answer our research question, we collaborated with an industrial partner from the automotive sector and developed a VA artifact capable of connecting, harmonizing, and visualizing measurements from a real-world manufacturing process for electrical engines. Following a Design Science Research (DSR) approach (Hevner et al. 2004; Vaishnavi and Kuechler 2008), we developed and instantiated the artifact design in close collaboration with our industrial partner. As a result, we propose the VA system ManEx (Manufacturing Explorer), which allows engineering experts to interactively analyze measurements from electrical engines across the manufacturing process using dedicated measures for anomaly detection. That being said, we contribute an artifact capable of supporting the analysis of large-scale IoT data, which is instantiated in a real-world manufacturing process for electrical engines and evaluated qualitatively and quantitatively with five engineering experts.

Related Work

In the context of serial manufacturing processes, global information networks emerge through large numbers of interconnected "Things". These networks comprise for example materials, sensors, or programmable logic controllers (Yang et al. 2019). This internet-based IoT infrastructure provides an unprecedented opportunity to connect different data generating instances from manufacturing processes to achieve the effective digital integration of the manufacturing stations. Recorded data can be used for several purposes, such as predictive maintenance (Dong et al. 2017) or energy efficiency management (Shaikh et al. 2017; Tan et al. 2017), to name a few. Due to the rapid growth of IoT sensing and the resulting big data challenges, new technologies, such as VA are needed to realize the full potential of IoT data in terms of data management and information processing (Yang et al. 2019). One way to address the challenges of IoT data is through visualization technologies. In this regard, VA aims to integrate the human into the visual exploration of complex data structures (Keim et al. 2008; van Wijk 2005) by improving analytical reasoning with model-supported and interactive interfaces (Thomas and Cook 2005). VA integrates humans into data analysis tasks with their perceptual abilities to get insights into the data, draw conclusions, and interact with the data through visual interfaces.

So far, VA systems have been applied to a wide range of IoT data in manufacturing contexts. For example, Cibuslki et al. (2020) provide a VA tool to show multi-criteria alternatives for the exploration of rotor designs. Xu et al. (2017) facilitate the exploration of IoT data for predictive maintenance in an automotive assembly line. Eirich et al. (2021) created a VA tool for the interactive analysis, clustering, and labeling of acoustic signatures of electrical engines. Suschnigg et al. (2020) use anomaly detection in multivariate time series data from test benches to detect faulty engines. Finally, Maier et al. (2012) propose a VA system that guides users in the detection of anomalies in a simulated manufacturing plant.

All aforementioned approaches contain the following key limitation: The presented VA applications use IoT data from only one source, such as single manufacturing stations (Eirich et al. 2021; Suschnigg et al. 2020), assembly line (Xu et al. 2017), or simulation (Maier et al. 2012; Post et al. 2017). However, using data from multiple sources enables analyzing the relation between different manufacturing stations across a manufacturing process and thus provides a bigger picture of the overall quality of produced parts. In this paper, we build on previous findings on designing and developing VA systems, especially for the analysis of large amounts of IoT data from manufacturing settings.

Methodology

This study followed a DSR approach addressing a significant but unsolved problem by designing and developing an artifact (Hevner et al. 2004) to analyze large amounts of IoT data in serial manufacturing processes. We used the five-phase DSR methodology proposed by Vaishnavi and Kuechler (2008) as demonstrated in Figure 1.

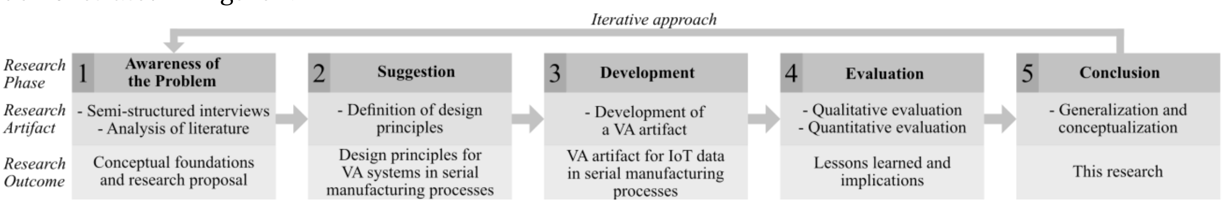


Figure 1. DSR Approach According to Vaishnavi and Kuechler (2008).

The study was carried out over nine months with a leading automotive manufacturer (hereafter "AutoCorp") based in Germany. Like many of its competitors, AutoCorp is investing considerable resources in the contextualization, exploration, and analysis of large-scale IoT data from manufacturing processes. We utilized the access to AutoCorp to conduct semi-structured interviews with engineering experts to get a better understanding of the problem domain, to install the artifact within AutoCorps IT landscape, and to evaluate our approach with engineering experts, who represent the core user group of *ManEx*. In particular, we instantiated and tested ManEx within a real-world setting with real IoT data from a serial manufacturing process for electrical engines.

For phase **(1) Awareness of the Problem,** we drew on existing literature on IoT data in the manufacturing industry and VA applications in the previous section. Furthermore, we carried out semi-structured interviews (Myers and Newman 2007) with six engineering experts working at AutoCorp. The interviews were held in person and lasted on average 33 minutes. Each interview followed a predefined guideline that covered questions about what problems engineers face when analyzing large amounts of IoT data and the potential uses of VA systems for supporting IoT analyses. One question for example was "*What challenges are you facing when you need to analyze measurements from two stations?*" The interviews were recorded, transcribed, and analyzed by the interviewer post hoc.

In the **(2) Suggestion** phase, we used the results from the literature (see Section "Related Work") and the interviews to develop nine design principles (DP), which represent normative design decisions that provide guidance on how an artifact should be instantiated (Gregor et al. 2020). Thus, the DPs build the initial foundation to instantiate our VA artifact capable of supporting the analysis and exploration of IoT measurements. To develop and specify the DPs, we followed the guidelines of Gregor et al. (2020).

In the **(3) Development** phase, we used the developed DPs to guide the instantiation of our DSR artifact in the form of the VA system *ManEx*. Following the idea of a user-centered and participatory design, this stage was performed in close collaboration with the same engineering experts as in phase **(1)**. We engaged with the experts in several workshops, gathered feedback, and tested design alternatives to continuously improve our artifact.

In the **(4) Evaluation,** we performed a qualitative and quantitative evaluation of our artifact with different engineers than the ones we developed the system with. The qualitative evaluation was performed following the *Think Aloud Method* of Someren et al. (1994), during which engineers were asked to express their thoughts during the use of a system. To quantitatively assess the usability and usefulness of our artifact, the engineers filled out the *System Usability Scale* (Brooke 1995).

In the **(5) Communication** phase, we concluded the project with the generalization and conceptualization of our results in this research.

Design Principles for a Manufacturing Exploration Artifact

Based on the interview results and the analyzed literature, we present the DPs for our VA artifact *ManEx* in Table 1. The conceptualization follows the three components of VA systems according to Keim et al. (2008). In this model, a VA artifact comprises a *data* (e.g., measurements), a *model* (e.g., anomaly detection in measurements), and a *visualization component* (e.g., visual representations of anomalies). To derive DPs, we used the guidelines from Gregor et al. (2020), where typical DPs comprise an *aim*, an *implementer*, a *user*, a *context*, a *mechanism*, and a *rationale*. All our DPs are embedded in the *context* of the analysis of IoT data in a serial manufacturing process for electrical engines, *implemented* by a VA expert, while the *users* are engineering experts. Since the *aims*, *mechanisms*, and *rationales* differ for each DP, we summarized them for the *data*, *model*, and *visualization component* in Table 1.

The **Data Component** comprises the collection of sensor data across a manufacturing process, which we hereafter refer to as measurements. Here, we derived the following DPs:

DP-D-1.1 Principle of Data Contextualization: When starting developing *ManEx*, we experienced that it was not clear how stations were related to each other. This, however, is crucial for the efficient tracking of parts and the identification of root causes of errors. Hence, we propose that as soon as measurements from multiple working stations are connected, a common semantic is needed to put them in context.

DP-D-1.2 Principle of Data Filtering: Connecting data from different sources results in a cohesive set of manufacturing data. The interviewed experts, however, reported that it is a cumbersome and time-consuming task to manually analyze such complex and large datasets. In this regard, one expert reported that "*It is hard to analyze big datasets manually because I have to apply the same filters every time, I perform an analysis.*" Thus, we propose filtering features, which experts can apply during system use.

DP-D-1.3 Principle of Interesting Parts: Our interviewed experts reported that one of their central tasks is to compare a group of parts, which contains an error to parts without errors. Since errors are often detected at rather late manufacturing steps, experts wanted to identify the root cause of the error as early as possible in the manufacturing process. Hence, it is necessary to provide mechanisms, which allow the comparison of measurements from groups of produced parts.

	Aim	*Mechanism*	*Rationale*
Data component	***DP-D-1.1 Principle of Data Contextualization***		
	Contextualize data from all distinct data sources	Use a common semantic to contextualize data	Providing a contextualized data set results in a big picture of the quality of produced parts
	DP-D-1.2 Principle of Data Filtering		
	Reduce the amount of data that must be analyzed	Include interactive filtering mechanisms	Removing unimportant information reduces analysis time
	DP-D-1.3 Principle of Parts of Special Interest		
	Compare groups of parts with an error to parts with no error	Enable the upload of parts of interest	Comparing two parts groups allows the early detection of faults
Model component	***DP-M-1.1 Principle of Data Harmonization***		
	Compare measurements from different sources	Compute residual values from the measurements	Computing residual values make measurements comparable
	DP-M-1.2 Principle of Hypothesis Testing		
	Identify measurements of interest for specific parts	Compute two-sided t-tests for groups of parts	Computing t-tests helps to identify interesting measurements
	DP-M-1.3 Principle of Focus-Set Selection and Validation		
	Build and validate a subset of measurements	Apply dimensionality reduction to the subset	More than three dimensions of measurements cannot be visually represented
Visualization component	***DP-V-1.1 Principle of the Hierarchical Representation***		
	Visualize the hierarchy of parts, steps, and stations	Use a tree representation	Tree views allow identifying errors for parts, steps, and stations
	DP-V-1.2 Principle of the Temporal Representation		
	Visualize the temporal dimension of part assembly	Use a directed graph	Directed graphs situate relations of parts, steps, and stations
	DP-V-1.3 Principle of Anomaly Abstraction		
	Enable the identification and assessment of anomalies	Provide different anomaly metrics and views	Different anomaly abstractions and metrics allow assessing anomalies and their causes

Table 1. Design Principles for our Artifact for the Data, Model, and Visualization Component.

The **Model Component** comprises routines to process available measurements of a manufacturing process and contains the following DPs:

DP-M-1.1 Principle of Data Harmonization: To compare measurement values the interviewed experts requested to harmonize them across the manufacturing process. For example, one expert noted, "*Since all measurements can be potentially interesting, we need an efficient way to compare them to each other across the manufacturing process.*" Thus, to show the deviation from an expected measurement to its desired result, we compute residual values. Here, for a measurement, we calculate the mean value as well as the standard deviation of a measurement from all parts. Next, for all parts, we subtract the original measurement of a part from the mean value of the measurement. The result is divided by the standard deviation, which results in the residual value for a measurement. We then aggregate the measurement residuals on a station level, next on a step level, and last on a part level.

DP-M-1.2 Principle of Hypothesis Testing: As soon as residuals are available, groups of parts can be compared. The interviewed experts suggested to "*use simple and easy to understand statistical measures to identify measurements, which are potential causes of an error.*" Thus, we propose to use hypotheses tests. Such tests ease the analysis of the difference in measurements between a group of parts with an error and a group of parts without errors.

DP-M-1.3 Principle of Variable Subset Selection and Validation: The experts reported that "*in many cases, a small portion of all recorded measurements is sufficient to detect an error*". Hence, it is necessary to add and validate measurements to a subset of measurements, which we call a *focus-set*.

The **Visualization Component** comprises interactive visualization interfaces, which enable the analysis of large amounts of measurements from a manufacturing process and contains the following DPs:

DP-V-1.1 Principle of the Hierarchical Representation: The interviewed experts stated that each produced part inherits fine-grained steps, each of which contains manufacturing stations that record measurements. To represent this hierarchy, we propose to visualize the hierarchical dimension of a manufacturing process as a tree structure, where a produced part forms the root and inherits steps, each of them inheriting stations, which contain measurements.

DP-V-1.2 Principle of the Temporal Representation: Parts that are assembled into new parts also change conditions over time. Thus, the interviewed experts requested to shed light on the temporal evolution of the part assembly and the relation of steps, stations, and measurements over time. Hence, we propose to use directed graphs, which are helpful to identify time-dependent relations.

DP-V-1.3 Principle of Anomaly Abstraction: The experts noted that it is important to differentiate between anomalies and statistical outliers, where one expert explained "*once I identified an anomaly, it is unclear if the anomaly is a statistical outlier or a new error*". To facilitate anomaly assessment, we propose metrics to compute anomalies, such as the mean value of all anomalies in a station.

Instantiation of the Manufacturing Explorer

Data component: The input data for *ManEx* is recorded at over 300 sensors at 154 stations along the manufacturing process of electrical engines. To connect IoT data from different sensors we use a semantic contextualization of the IoT measurements, which are embedded into AutoCorps IT infrastructure. This semantic contextualization contains information on how stations are related to each other, which measurements they include, and to which manufacturing step they belong. With the semantic available, *ManEx* aggregates the data from different stations, steps, and measurements into a new cohesive dataset, which forms the basis for any analysis scenario (*Principle of Data Contextualization*). Next, engineering experts can judge, if a data instance is important or not. Thus, we include an interactive filtering option into the visualization component to enable engineering experts to filter datasets (*Principle of Data Filtering*). To compare different groups of produced parts (*Principle of Parts of Special Interest*), engineering experts can upload a set of part IDs into *ManEx* or directly select engines of interest in *ManEx* itself. *ManEx* then produces a separate comparison dataset, which includes only measurements that can be related to the uploaded part IDs. Experts can then compare a group of erroneous parts to parts with no error and identify the root cause of an error in the manufacturing process.

Model component: Each station records heterogeneous IoT measurements (e.g., temperature). To make recorded measurements comparable (e.g., voltage vs. pressure), we follow approaches of similar VA systems in the manufacturing sector (Suschnigg et al. 2020) and the suggestions of our interviewed engineering experts to compute the residual values for each measurement instead of using absolute values (*Principle of Data Harmonization*). As a result, the deviation from expected measurement values stands out, which eases the identification of errors. To identify different measurements that show significant differences in the recorded data from two distinct part groups, *ManEx* contains a hypothesis module (*Principle of Hypothesis Testing*). In this module, engineering experts can upload part ids that they identified as erroneous in previous manual analyses. Next, the system computes a two-sided t-test for the measurements from the erroneous parts to a control group of parts with no error and suggests measurements, where two samples are significantly different according to the p-values. To build and validate a subset of interesting measurements (*Principle of Focus-Set Selection and Validation*), engineering experts can sequentially add and delete measurements to their *focus-set*. As soon as more than three measurements are added to the *focus-set*, *ManEx* computes a dimensionality reduction to project linear dependencies in the dataset to a two-dimensional space. We use a principal component analysis (PCA) as a dimensionality reduction algorithm, which is a well-established method to identify linear relationships in high-dimensional data (Subasi and Ismail Gursoy 2010).

Visualization Design

The final visualization design of *ManEx* contains five views (A-E) and is displayed in Figure 2. In (A), an overview of the hierarchical dimension of the manufacturing process is displayed as a tree structure. Selecting a part unfolds its steps and selecting a step further enables a drill down to its stations (*Principle of the Hierarchical Representation*). The temporal relation of the manufacturing process is shown as a directed graph for parts and steps (*Principle of the Temporal Representation*). Two manually added gray arrows indicate the hierarchical and temporal dimensions of the manufacturing process in (A). Please note that stations all run in parallel and thus are not displayed as a directed graph. To support the differentiation of part groups across all views, each group contains qualitative colors (e.g., orange and blue in Figure 2), while the control group is always displayed in gray. Anomaly scores are displayed as two-colored glyphs. The arc size depends on their residual values, where bigger residual values result in bigger arcs. Blue glyphs represent negative and red glyphs positive residual values from the measurements (*Principle of Anomaly Abstraction*). Furthermore, users can choose between different metrics to calculate anomaly scores, such as the mean value of all anomalies, the maximum of all anomalies, or the median of all anomalies.

T-test results for part groups are displayed in (B) and ordered according to their p-value. Users can add a new part group to *ManEx* in (A) (see the red outline at the top in Figure 2-A). By doing so, the t-tests are recalculated and all views are adapted accordingly. By selecting a station in (A) or a row from (B), measurements are shown in (C) as a boxplot view. If too many measurements are displayed in (C), users can apply filters in (D) to reduce the number of available boxplots. Each measurement contains a boxplot for all part groups, while different measurements are separated with gray lines. By selecting a boxplot of interest, users can add a measurement to the *focus-set* (see the red outline in Figure 2-B). The PCA representation of the *focus-set* is displayed in (E) as a scatterplot. When users add or delete measurements to and from the *focus-set*, the PCA is recalculated and (E) adapted accordingly.

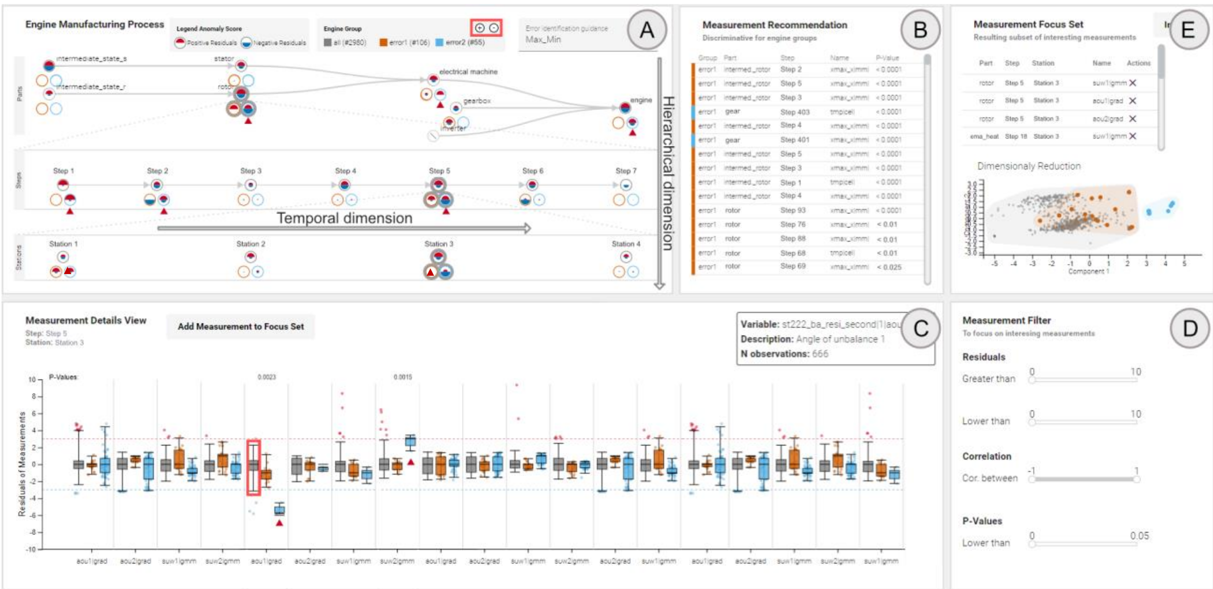


Figure 2: Views of the ManEx System to Analyze IoT Data from a Manufacturing Process.

These visualizations can now be used by engineering experts inside AutoCorp to derive insights out of IoT data. Figure 2 shows three groups of electrical engines, where the erroneous engines were detected by a test bench located at the last step of the manufacturing process of engines. Manually added red triangles in (A) and (C) indicate that for the blue error group ("*error 2*" in the legend of Figure 2-A) anomalies exist on a part level in the engine, the electrical machine, and the rotor. On a step level, steps 1, 2, and 5 contain high anomalies. For step 5, the highest anomalies were produced at stations 1 and 3. By inspecting the boxplots of station 3, we can identify the two measurements, which deviate from the control group of gray engines (see red triangles in Figure 2-C). In the example, an engineering expert added 23 measurements to his *focus-set,* which resulted in the PCA representation in (E). Again, we can see that the blue engines are visually different from all other engines. This indicates that the 23 selected measurements serve well to detect the error from the blue engine group at manufacturing steps. Engineers could now use the 23 measurements to build a new test procedure to prevent this error and thus improve the overall quality of the manufacturing process.

Evaluation

Our evaluation aims at validating the quality in terms of usability and usefulness of our presented VA artifact *ManEx*. We evaluate the artifact qualitatively and quantitatively. For the qualitative evaluation, we performed a think-aloud study (Someren et al. 1994) with one observer taking notes. First, we gave a detailed system walk-through, where participants were able to ask questions. To test the artifact with our industrial partner, we uploaded 2980 electrical engines as a control group and one group of 106 erroneous engines. All engines contained 26,160,540 measurements recordings, which were produced at 54 steps and 154 stations. The data was recorded during a time of six weeks of production at AutoCorp. Next, we asked all participants to perform the following task: *"Please find a set of measurements, which you believe is relevant to develop a testing procedure for the specific error",* followed by open-ended questions on how each view of *ManEx* supported the execution of the task. All interviews were held in person and lasted on average 90 minutes. We further applied the *System Usability Scale* (SUS) (Brooke 1995) after each think-aloud session. The SUS contains ten questions (e.g., I felt very confident using the system), which participants had to rate on a five-point Likert scale.

Qualitative Results from the Think-Aloud Study

In terms of the qualitative evaluation, we observed that the engineers performed the same workflow using *ManEx*. First, we provided all engineers with an exemplary dataset of erroneous engines (*DP-D-1.3*), which they uploaded to *ManEx* to find out the root cause of the error. Next, they navigated through the hierarchical (*DP-V-1.1*) and temporal (*DP-V-1.2*) dimensions of the manufacturing process using the anomaly glyph representation (*DP-V-1.3*) to compare the erroneous engines to the control group of error-free engines. They then selected the station containing the highest anomalies and added measurements from the station to their *focus sets*. One engineer explained his decision to add a measurement to the *focus-set* with "*based on the measurement name I can tell if it makes sense to add this measurement to my focus-set or not*". Since the station contained over 80 measurements, the participants applied filters to display only measurements with high positive deviations (*DP-D-1.2*). In this regard, one engineer noted "*I am only interested in high positive residuals because these are good indicators for the error I am analyzing*". Next, the engineers used the t-test table (*DP-M-1.2*) to find measurements of more stations, where the erroneous and error-free engines deviated from each other. Regarding the t-test, an engineer acknowledged that "*the t-tests provide an additional help to identify interesting measurements, which I did not find with the anomaly glyph representation"*. While adding and deleting measurements in the *focus sets*, the engineers constantly reviewed the PCA visualization until they were satisfied with the PCA scatterplots (*DP-M-1.3*). An engineer stated that it is "*very valuable to immediately see how my focus-set is capable of splitting the good engines from the bad ones*." While using *ManEx*, the participants reported that "*it is very important to analyze deviations instead of absolute values to compare different types of measurements*" (*DP-M-1.1*). Furthermore, all participants recognized that *ManEx* is the first VA system at AutoCorp capable of connecting, contextualizing (*DP-D-1.1*), and visualizing data from an entire manufacturing process. Since some measurement names were not self-explanatory, one engineer suggested including detailed descriptions of the meaning of measurements in *ManEx* but stated "*It is a little bit unfair to request such features from ManEx because actually, we have to provide this information ourselves when we define the measurement names at our stations*". Apart from that, the engineers did not suggest further system improvement.

Quantitative Results from the System Usability Scale

Regarding the SUS, we found that our systems' usability with a rating of 83.5 is high above the average score of 68 (Lewis and Sauro 2018). The individual scores for each question (Q) are outlined in Table 2. All questions of the *System Usability Scale* can be found in Lewis and Sauro (2018). Even though we evaluated the system with only five engineers, we were confident that our system reached a sufficient level of usability. Apart from the evaluation results, engineers are currently using *ManEx* to support error detection in the manufacturing process and improve product quality. This already resulted in a reduction in scraped parts of over 3%. In terms of serial manufacturing processes producing thousands of engines each day, the saved costs would go into the millions per year.

	Q1	*Q2*	*Q3*	*Q4*	*Q5*	*Q6*	*Q7*	*Q8*	*Q9*	*Q10*	*Total*
Expert 1	10	7.5	7.5	10	7.5	7.5	10	5.0	7.5	10	82.5
Expert 2	7.5	7.5	7.5	10	7.5	10	5.0	7.5	7.5	10	80.0
Expert 3	7.5	7.5	7.5	10	10	10	7.5	7.5	7.5	10	85.0
Expert 4	10	10	10	10	10	7.5	10	7.5	7.5	5.0	87.5
Expert 5	7.5	7.5	10	7.5	7.5	10	7.5	7.5	10	7.5	82.5
Avgerage	8.5	8.0	8.5	9.5	8.5	9.0	8.0	7.0	8.0	8.5	83.5

Table 2. Results from the System Usability Scale with Five Engineering Experts.

Discussion

To address our research question on how to design a VA artifact for large-scale IoT manufacturing data, we used a DSR approach, which outlined how we developed and instantiated the VA system *ManEx*. The overall success story of our artifact is demonstrated by the fact that study participants perceived *ManEx* as highly usable and useful. Furthermore, the artifact showed great potential to reduce scrapped parts and improve the overall quality of AutoCorp's manufacturing processes. Reflecting on the development of *ManEx*, we can translate some of our lessons learned into more general guidelines. These guidelines can help researchers and practitioners to respond to the challenges arising from the analysis of IoT data from large-scale manufacturing processes and to develop similar VA artifacts.

Data Component: To analyze IoT data along a manufacturing process, large datasets are recorded at heterogeneous stations. This challenge was not addressed by previous VA approaches in the manufacturing industry. One stream of research included the analysis of very detailed data spaces from isolated real-world manufacturing data sources (Eirich et al. 2021; Suschnigg et al. 2020). In this regard, Eirich et al. (2021) noted that the analyzed data contained a high level of detail but that it is necessary to put sensor data into context with other manufacturing stations. The other stream analyzed a broader spectrum of data sources but only included synthetic datasets (Post et al. 2017). Although we did not provide a very fine-grained level of detail in the data, our goal was to put the data into context with different steps along a manufacturing process. The focus was to make parts traceable along a manufacturing process and to identify root causes of detected errors. We encountered several challenges, such as absolute measurement values that could not be compared or IoT data sources that needed to be contextualized first to understand their relation. These challenges were addressed through the three DPs of *data contextualization (DP-D-1.1)*, *data filtering (DP-D-1.2)*, and *parts of special interest (DP-D-1.3).* We believe that these principles can guide researchers and practitioners to avoid challenges that arise when developing artifacts for the analysis of manufacturing processes.

Model Component: To make heterogeneous measurements comparable, we propose to use residual values to make data from different locations of the manufacturing process comparable. The benefit of residuals is that statistical outliers, which are strong indicators of anomalies (Suschnigg et al. 2020), can be identified easily by human experts. When dealing with high-stakes decision-making processes, often sophisticated ML approaches are used to identify patterns in the data (Eirich et al. 2021; Suschnigg et al. 2021). Many of the resulting ML models, however, are black boxes not capable of explaining their predictions in a way that humans can understand. This lack of transparency can have severe consequences, such as a lack of trust in the system or even human experts refusing to use the system (Rudin 2019). In turn, we propose to find measurements of interest by using t-tests and validate a set of interesting measurements using PCA. These are white box, explainable, and well-established yet simple statistical measures, which the engineering experts reported to be especially helpful in the identification of root causes for errors in the manufacturing process. Hence, we hope that our DPs for *data harmonization (DP-M-1.1), hypothesis testing (DP-M-1.2)*, and *focus-set selection and validation (DP-M-1.3)* guide the effective application of white box modeling approaches to real-world problems while including human experts into the loop of data analysis tasks.

Visualization Component: All industrial manufacturing processes have in common that they contain both a hierarchical and temporal dimension. Researchers and practitioners should take these two perspectives into account when designing visualization interfaces for similar domain problems. In our artifact, we showed that tree representations and directed graphs serve well to represent this two-dimensional space. Furthermore, our proposed two-colored glyph visualization proved to be an adequate representation to easily identify anomalies. Hence, we propose that the principles of the *hierarchical representation (DP-V-1.1)*, the *temporal representation (DP-V-1.2)*, and the *anomaly abstraction (DP-V-1.3)*, fit well to easily navigate along with manufacturing processes, locate errors, and understand the root cause of detected errors.

Compared to other VA systems, such as ViDX (Xu et al. 2017) that support the analysis of manufacturing IoT data, we believe that ManEx provides some significant advantages. For example, ViDX only uses timestamps and error codes as measurements, which is visualized as a Marye's graph. In turn, ManEx focuses more on numerical measurements (e.g., temperature or voltages), where we process thousands of distinct measurements. Another example is that ViDX particularly targets the temporal dimension of the manufacturing process taking only stations and steps into account. ManEx, however, provides a different contextualization of the manufacturing process taking the hierarchical dimension also into account including produced parts as an additional layer of the manufacturing process.

We acknowledge that our study has some limitations. The artifact was instantiated in the IT landscape of only one company in the automotive industry. However, we believe that our DPs have a sufficient level of abstraction to affect the generalizability of building VA artifacts in other manufacturing-related systems. For example, one could consider the aviation industry, where plane turbines are produced along a hierarchical and temporal dimension and IoT data is recorded in distinct sensors. The interpretation of our results is also constrained by the small size of the evaluation. Because VA systems are often designed to tackle very specific domain problems, they naturally address relatively small user groups as demonstrated by similar VA approaches (Cibulski et al. 2020; Eirich et al. 2021; Suschnigg et al. 2020). Thus, our results do not allow for drawing any conclusion regarding the empirical validity of our DPs in our application domain.

Conclusion

In conclusion, we present a DSR approach on the development of a VA artifact to analyze large numbers of IoT data from different stations of a manufacturing process. The resulting VA artifact *ManEx* provides a clear picture of the hierarchical and temporal dimensions of the manufacturing process. It further supports engineering experts with the tracking of the quality of parts, the identification of errors, and the analysis of root causes of detected errors. Furthermore, we outline nine DPs, which indicate how a VA artifact needs to be designed to support engineering experts with the analysis of IoT data. We instantiated the artifact within the IT landscape of our industrial partner, where it is currently used to optimize the quality of producing electrical engines. Finally, we demonstrate the usability and usefulness of *ManEx* with both a qualitative and quantitative evaluation with five engineering experts, in line with the reduction of scrapped parts in a real-world manufacturing scenario. Fueled by the positive feedback of our industrial partner, we will further develop the artifact in future research projects. For example, we intend to include more experts to provide further quantitative evidence of the usefulness of our approach.

REFERENCES

Brooke, J. 1995. "SUS: A Quick and Dirty Usability Scale," *Usability Eval. Ind.* (189).

Cibulski, L., Mitterhofer, H., May, T., and Kohlhammer, J. 2020. "PAVED: Pareto Front Visualization for Engineering Design," *Computer Graphics Forum* (39), pp. 405–416.

Dong, L., Mingyue, R., and Guoying, M. 2017. "Application of Internet of Things Technology on Predictive Maintenance System of Coal Equipment," *Procedia Engineering* (174), pp. 885–889.

Eirich, J., Bonart, J., Jäckle, D., Sedlmair, M., Schmid, U., Fischbach, K., Schreck, T., and Bernard, J. 2021. "IRVINE: A Design Study on Analyzing Correlation Patterns of Electrical Engines," *IEEE Transactions on Visualization and Computer Graphics*.

Gregor, S., Kruse, L., and Seidel, S. 2020. "Research Perspectives: The Anatomy of a Design Principle," *Journal of the Association for Information Systems* (21), pp. 1622–1652.

Hevner, March, Park, and Ram. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), p. 75.

Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H. 2008. "Visual Analytics: Scope and Challenges," in *Visual Data Mining* (Vol. 4404), S. J. Simoff, M. H. Böhlen, and A. Mazeika (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 76–90.

Kusiak, A. 2018. "Smart Manufacturing," *International Journal of Production Research* (56:1–2), pp. 508–517.

Lewis, J. R., and Sauro, J. 2018. "Item Benchmarks for the System Usability Scale," *J. Usability Studies* (13:3), Bloomingdale, IL: Usability Professionals' Association, pp. 158–167.

Maier, A., Tack, T., and Niggemann, O. 2012. "Visual Anomaly Detection in Production Plants:," in *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics*, Rome, Italy: SciTePress - Science and and Technology Publications, pp. 67–75.

Myers, M. D., and Newman, M. 2007. "The Qualitative Interview in IS Research: Examining the Craft," *Information and Organization* (17:1), pp. 2–26.

Post, T., Ilsen, R., Hamann, B., Hagen, H., and Aurich, J. C. 2017. "User-Guided Visual Analysis of Cyber-Physical Production Systems," *Journal of Computing and Information Science in Engineering* (17:2), p. 021005.

Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence* (1:5), pp. 206–215.

Shaikh, F. K., Zeadally, S., and Exposito, E. 2017. "Enabling Technologies for Green Internet of Things," *IEEE Systems Journal* (11:2), pp. 983–994.

Someren, M. W. van, Barnard, Y. F., and Sandberg, J. A. C. 1994. *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*, London: Academic Press.

Subasi, A., and Ismail Gursoy, M. 2010. "EEG Signal Classification Using PCA, ICA, LDA and Support Vector Machines," *Expert Systems with Applications* (37:12), pp. 8659–8666.

Suschnigg, J., Mutlu, B., Fuchs, A. K., Sabol, V., Thalmann, S., and Schreck, T. 2020. "Exploration of Anomalies in Cyclic Multivariate Industrial Time Series Data for Condition Monitoring," in *Proceedings of the EDBT/ICDT 2020 Joint Conference Workshops*, CEUR Workshop Proceedings.

Suschnigg, J., Mutlu, B., Koutroulis, G., Sabol, V., Thalmann, S., and Schreck, T. 2021. "Visual Exploration of Anomalies in Cyclic Time Series Data with Matrix and Glyph Representations," *Big Data Research* (26), p. 100251.

Tan, Y. S., Ng, Y. T., and Low, J. S. C. 2017. "Internet-of-Things Enabled Real-Time Monitoring of Energy Efficiency on Manufacturing Shop Floors," *Procedia CIRP* (61), pp. 376–381.

Thomas, J. J., and Cook, K. A. (eds.). 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, Los Alamitos, Calif: IEEE Computer Soc.

Vaishnavi, V., and Kuechler, W. 2008. *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, Boca Raton: Auerbach Publications.

van Wijk, J. J. 2005. "The Value of Visualization," in *VIS 05. IEEE Visualization, 2005.*, Minneapolis, MN, USA: IEEE, pp. 79–86.

Xu, P., Mei, H., Ren, L., and Chen, W. 2017. "ViDX: Visual Diagnostics of Assembly Line Performance in Smart Factories," *IEEE Transactions on Visualization and Computer Graphics* (23:1), pp. 291–300.

Yang, H., Kumara, S., Bukkapatnam, S. T. S., and Tsung, F. 2019. "The Internet of Things for Smart Manufacturing: A Review," *IISE Transactions* (51:11), pp. 1190–1216.

6.9 P9: Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms

Full reference: *J. Bonart, J. Eirich, P. Hümmer, and W.-G. Drossel. Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms. In IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, 08 2021*

Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms.

J. A. E. Bonart, J. Eirich, P. Huemmer, W.-G. Drossel

***Abstract*—Series production of electric drives tends towards highly-integrated topologies. All components ranging from the inverter over the electric engine to the reduction gear are mounted in a single housing. The identification of faults by measuring structure-borne noise exhibits new challenges as each vibration of a part is dependent on its surroundings. A metric for calculating a measure for anomaly regarding the mean vibration signature is compared to known metrics. The reduced difference-correlation metric is exploited further to address anomalies across different order tracks, consequently scanning spectrograms for anomalous coincidences. The method is evaluated by two classes of faults (electrical and mechanical). The faults can not be distinguished by elementary order analyses as they match in their primary excitation. Including the view of secondary fault excitations each fault can be classified. The interconnection between different emerging orders is clearly observable and the faults can be classified.**

***Index Terms*—Acoustic testing, Statistical analysis, Harmonic Analysis, Electric machines, Gears, Vibration measurement**

I. Nomenclature

EOL	End-of-Line
rpm	Revolutions per minute
$\mathcal{O}$	Engine order
n	Engine speed
μ	Mean value
σ	Standard deviation
z_i	Number of teeth on the $i-$th gear
N	Number of balls in a bearing
d, D	Ball and pitch circle diameter

II. Introduction

The development of the next generation of electric drives heads towards highly-integrated topologies. This trend allows for higher power densities and lower manufacturing cost as the housing is shared for all components. The increased compactness introduces new interactions between rotating parts as each deviation from a normal movement (primary excitation) excites other parts which are near the source (secondary excitation). The series production of electric drives consists of a line of production with different variants of units depending on their position in the vehicle. Therefore a transferable approach for all different variants is needed. Simulations of interactions and co-excitations are reported by [1], [2]. The identification of different faults by black-box approaches is the focus of [3]–[8]. These approaches reach a precision of approx. 98% for laboratory setups. None of the approaches, however, is directly transferable to different variants of the same topology, since black-box approaches are each optimized for the data at hand. White-Box approaches mainly approach a dedicated fault by different transformations like order-FFT [9] or envelope spectra [10]–[12]. In order to identify their fault additional filtering [13] or demodulation techniques [14] are proven to be useful for modulated signals and their resulting sideband phenomena in a normal FFT. The repetition of equidistant peaks inside a spectrum can be examined by cepstrum techniques as shown by [15]. The main downside of this approach is the need for periodic patterns which does not address sideband phenomena accordingly. While each paper is able to isolate their single fault with good precision (the authors do not report percentages as they mainly compare a single healthy to a faulty unit). None of the methods proposes a distinct way to distinguish between different faults which primarily excite the same engine order. Accordingly, to unveil secondary mechanisms and classify them according to expert knowledge, a unified white-box approach to identify anomalies is needed.

The paper is structured as follows. Firstly the test environment, consisting of the electric drive topology and the signal preprocessing, is described in Sec. III. The different metrics which build on the foundation of the preprocessing and a comparison between them is given in Sec. IV. The reduced difference correlation metric is employed further to identify secondary excitation mechanisms for two classes of faults (electrical and mechanical). The latter consists of three different examples, each primarily exciting the order corresponding to the outer ring of a bearing.

III. Test environment

In this section, the electric drives topology and signal preprocessing are described to lay a foundation for the methods described latter.

A. Electric drive topology

The electric drives are constructed as highly-integrated units. Thus, the electric drive consisting of an inverter, stator,

J. A. E. Bonart is with BMW Group, Taunusstrasse 41, 80807 Munich, GERMANY (e-mail: jakob.bonart@bmw.de) and Fraunhofer IWU Dresden, Nöthnitzer Straße 44, 01187 Dresden GERMANY.

J. Eirich and P. Huemmer are with BMW Group, Taunusstrasse 41, 80807 Munich, GERMANY.

W.-G. Drossel is with Fraunhofer IWU Dresden, Nöthnitzer Straße 44, 01187 Dresden GERMANY.

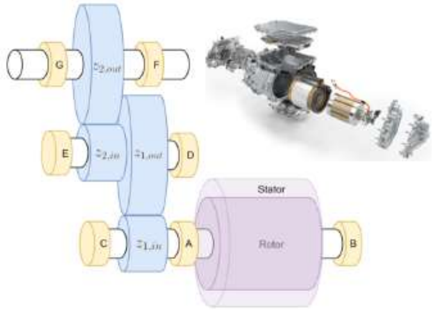


Fig. 1. Schematic topology of the electric drives which are used to validate the methods proposed in this paper. Gears are colored blue, shafts are displayed as white. Bearings are shown yellow and the electromagnetic components are violet.

rotor, a two-stage spur gear and all necessary bearings are located in a compact housing. A general overview of the topology is given in figure 1. The electric drive is classified as an electrically excited synchronous machine. The inverter is placed on top into the housing and supplies both stator and rotor with currents. The rotor's bearings are called A- and B-bearing. The connection between the rotor and the gearbox input shaft is realized via an involute spline. The two gear stages reduce the rotation speed to the desired speed of the tires. The number of teeth at the input shaft is denoted by $z_{1,\mathrm{in}}$, the corresponding number of teeth at the intermediate shaft are called $z_{1,\mathrm{out}}$. The second stage is named $z_{2,\mathrm{x}}$, accordingly. The associated output shaft is supported by the F and G bearing. As all tests were carried out in the EOL test bench of the series production the tires are replaced by load machines on each side. During the test, the electric drive provides an asked amount of torque corresponding to the real driving situation.

B. Signal preprocessing

To lay the foundation for the anomaly detection process, the necessary preprocessing steps are outlined in the following paragraph. The result of preprocessing is an order spectrogram. Each step of preprocessing is done by a Tasalyzer system from Discom [16]. The vibration sensor is applied at the gear housing and measures unidirectional at the position of the C-bearing along the shaft axis of the rotor (compare to figure 1). The raw signals from the EOL are measured with a sampling rate of 100 kHz. The broad frequency range is chosen to maintain a safe distance to the Nyquist-Shannon frequency and corresponding problems like aliasing. The angle of the drive shaft is captured by 360 pulses per revolution. It is then transferred towards an angle at the rotor shaft regarding the gear ratio. The resampling of the vibration signals from the time-domain towards the angular-domain ensures minimal smearing effects in the following FFT-results. The FFT is carried out over 16 revolutions with a resolution of 4096 points. A von-Hann window is used to comply with the demand for periodicity. The Fourier-transformation of a signal in the angular-domain is called order-spectra. An engine order $\mathcal{O}$ corresponds to an excited frequency $f = \mathcal{O} \times n$ for a given speed n. The order-spectra are ordered along the rising speed resulting in an order-spectrogram. The x-axis describes the order and the y-axis displays the corresponding speed. In our case all engine orders are referred to the shaft speed of the rotor. The engine orders of gear meshing are located at a single order bin (e.g. 23 teeth on the input shaft $\longrightarrow$ 23th order of the input shaft). The amplitudes of an order along a speed ramp are called order track. We chose standard preprocessing steps allowing for lean transfers to other research groups if they want to evaluate our approach with their own data. Exotic windowing functions or order resampling techniques may inhibit an easy testing as non-essential routines have to be implemented as well. An overview of orders excited during a ramp and their source is listed in tab. I. The system exhibits three rotation speeds of shafts $\mathcal{O}_{\mathrm{S}}$. They are related by the number of teeth z_i at the stages of the spur gear. The order of each component is scaled by the corresponding shaft accordingly. The orders of the bearing are calculated by the diameter of the balls d, the number of balls N, the pitch circle diameter D and the pressure angle α. Based upon the elementary orders from the bearing, the orders excited by a waviness Z at the rings can be calculated, too.

Each of the orders displayed in tab. I is to be seen as the elementary order. As none of the excitation mechanisms can be classified as a pure sinusoidal, it is obvious that harmonics of the elementary order are excited, too.

TABLE I
OVERVIEW OF ORDERS AND THEIR CORRESPONDING SOURCE.

Source	[1], [17]–[21]		Order $\mathcal{O}$
Unbalance	Rotor shaft	$\mathcal{O}_{\mathrm{S-R}}$	reference order
	Intermed. shaft	$\mathcal{O}_{\mathrm{S-I}}$	$\mathcal{O}_{\mathrm{S-R}}\frac{z_{1,\mathrm{in}}}{z_{1,\mathrm{out}}}$
	Output shaft	$\mathcal{O}_{\mathrm{S-O}}$	$\mathcal{O}_{\mathrm{S-R}}\frac{z_{1,\mathrm{in}}z_{2,\mathrm{in}}}{z_{1,\mathrm{out}}z_{2,\mathrm{out}}}$
Bearing	Cage rotation	$\mathcal{O}_{\mathrm{C}}$	$\frac{\mathcal{O}_{\mathrm{S}}}{2}\left(1-\frac{d}{D}\cos(\alpha)\right)$
	Outer ring	$\mathcal{O}_{\mathrm{bpo}}$	$\frac{N\mathcal{O}_{\mathrm{S}}}{2}\left(1-\frac{d}{D}\cos(\alpha)\right)$
	Inner ring	$\mathcal{O}_{\mathrm{bpi}}$	$\frac{N\mathcal{O}_{\mathrm{S}}}{2}\left(1+\frac{d}{D}\cos(\alpha)\right)$
	Ball rotation	$\mathcal{O}_{\mathrm{b}}$	$\frac{D\mathcal{O}_{\mathrm{S}}}{2d}\left[1-\left(\frac{d}{D}\cos(\alpha)\right)^2\right]$
Waviness	Outer ring	$\mathcal{O}_{\mathrm{wo}}$	$\mathcal{O}_{\mathrm{S}}$ and $Z\mathcal{O}_{\mathrm{C}}$
	Inner ring	$\mathcal{O}_{\mathrm{wi}}$	$Z\left(\mathcal{O}_{\mathrm{S}}-\mathcal{O}_{\mathrm{C}}\right)$
Magnetic	pole number	$\mathcal{O}_{\mathrm{em}}$	6
	stator slots	$\mathcal{O}_{\mathrm{stator}}$	54
Gear	fundamental	$\mathcal{O}_{\mathrm{gm}}$	$\mathcal{O}_{\mathrm{S}}z_i$
	eccentricity	$\mathcal{O}_{\mathrm{gm,ecc}}$	$\mathcal{O}_{\mathrm{gm}} \pm \mathcal{O}_{\mathrm{S}}$

IV. DISTANCE METRICS

In this section, we discuss a range of metrics applicable towards the detection of anomalies in order tracks. Firstly two common metrices are described in order to set a baseline. In section IV-C the proposed approach is then presented and in section IV-D compared against the baseline metrices. Each metric refers to statistical quantities of the ensemble of tested

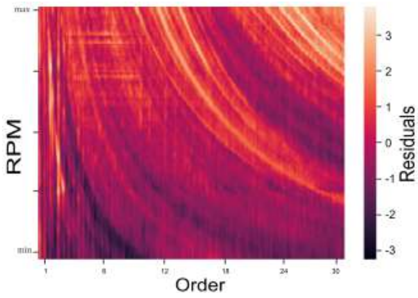


Fig. 2. Residual values of an exemplary order spectrogram.

electrical drives. The mean value of amplitudes for a given order $\mathcal{O}$ and speed n is denoted by $\mu_n^{\mathcal{O}}$. Consequently, the standard deviation follows the notation by $\sigma_n^{\mathcal{O}}$. The value of an electric drive unit is called $\nu_n^{\mathcal{O}}$.

A. *Euclidean metric*

An elementary metric for extracting the difference between a measurement and a reference, namely the mean of all drives measured, is the Euclidean metric. Each speed is assigned with an independent dimension. The measure of an anomaly is the square-root sum of all deviations from the mean value.

$$\Delta_{\text{Euclidean}}^{\mathcal{O}} = \sqrt{\sum_n \left(\nu_n^{\mathcal{O}} - \mu_n^{\mathcal{O}}\right)^2} \tag{1}$$

The simplicity of the Euclidean metric makes it a common choice for a broad range of applications. A downside of the Euclidean metric is its independence of any scatter parameters, namely ignoring available facts about typical deviations.

B. *Residual metric*

The residual metric consists of rescaling the space of deviations from the mean by the associated standard deviation:

$$\delta_{\text{Residual},n}^{\mathcal{O}} = \frac{\nu_n^{\mathcal{O}} - \mu_n^{\mathcal{O}}}{\sigma_n^{\mathcal{O}}} \tag{2}$$

Therefore, the resulting residual space is unitless. This allows for comparisons between processes that typically do not share the same unit or scale. Deviations that lie outside the typical scatter of measured values are also weighted more heavily than if they occur in regions with a high scatter. A typical residual spectrogram is shown in figure 2. Regions with values between $\pm 2[\sigma]$ can be interpreted as normal fluctuations. If the values rise above $\pm 3[\sigma]$ a high deviation compared to the normal scatter occurred. The summarized anomaly of an order track is obtained by the sum of absolute residual deviations over the whole speed interval which was measured.

$$\Delta_{\text{Residual}}^{\mathcal{O}} = \sum_n \left|\delta_{\text{Residual},n}^{\mathcal{O}}\right| \tag{3}$$

C. *Difference-correlation metric*

Both metrics shown beforehand rearrange an order track to a d-dimensional vector and compare each dimension by itself to the mean. Therefore, its is not possible to account for any interactions between the dimensions, nor to gain insights about them. Thus, we propose a two-point measure of an anomaly in order to identify anomalous dependencies in-between an order track. The advantage of a comparing metric is also the emerging possibility of tracking anomalous dependencies in-between different order tracks with a white-box approach, consequently unveiling secondary excitations, which correspond to a primary source.

The calculation of the difference-correlation metric for two order tracks is shown in figure 3. We start with an extraction of the order tracks A and B from the residual spectrogram for each electric drive in the ensemble. The coincidence matrix for a unit is given by:

$$\delta_{\text{d-c},n,n'}^{\mathcal{O},\mathcal{O}'} = \frac{\nu_n^{\mathcal{O}} - \mu_n^{\mathcal{O}}}{\sigma_n^{\mathcal{O}}} \otimes \frac{\nu_{n'}^{\mathcal{O}'} - \mu_{n'}^{\mathcal{O}'}}{\sigma_{n'}^{\mathcal{O}'}}. \tag{4}$$

The correlation matrix $\mathcal{M}^{\mathcal{O},\mathcal{O}'}$ of the pair of orders $\mathcal{O},\mathcal{O}'$ is consequently obtained by averaging all coincidence matrices over the whole ensemble of units.

$$\mathcal{M}_{n,n'}^{\mathcal{O},\mathcal{O}'} = \frac{1}{N} \sum_{\text{Ensemble}} \delta_{\text{d-c},n,n'}^{\mathcal{O},\mathcal{O}'} \tag{5}$$

The resulting difference-correlation metric for a pair of orders is calculated by substracting the correlation from the coincidence of a single unit. Consequently, this results in anomalous coincidences, which cannot be explained by normal correlation processes. If all values from a measured unit exhibit the same tendencies as the mean correlation for the pair of orders, the difference-correlation metric would be zero for all speeds. Processes that suppress normal correlations or reverse their effects, result in a negative difference of correlations and are detected accordingly. The rms-summation of all entries of a matrix is called Frobenius norm and used to boil down the values to a single value for the anomaly measure.

$$\Delta_{\text{d-c}}^{\mathcal{O},\mathcal{O}'} = \sqrt{\sum_{n,n'} \left(\delta_{\text{d-c},n,n'}^{\mathcal{O},\mathcal{O}'} - \mathcal{M}_{n,n'}^{\mathcal{O},\mathcal{O}'}\right)^2} \tag{6}$$

As the dimensionality and the information about interdependencies across different speeds is condensed to a single value, we call it a *reduced difference-correlation metric* (*rdc metric* or *rdcm*). Sorting all rdcm-values according to their pair of orders into a 2d array, one results in a hypermatrix. It displays the interaction of a single engine order with other engine orders, fundamentally depicting the source of an anomaly and secondary orders that are also excited by the irregularities in movement from the main excitation. Without the substraction of the mean correlation one would end up with strongly biased results depending on the normal interaction. As the aim of a metric for anomaly detection is to reduce normal signals, it is natural to substract the mean correlation matrix.

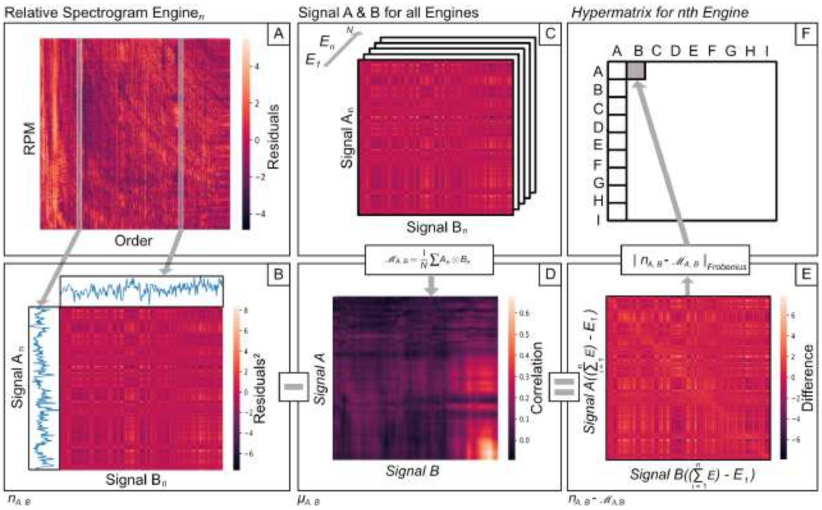


Fig. 3. Schematic description of the calculation of the reduced difference-correlation metric and the rearrangement into a hypermatrix.

D. Comparison of metrics

The comparison of the three metrics is undertaken by normalizing each by their mean value. The ensemble consists of 600 engines measured on the test bench. The anomaly measure for each metric is displayed in figure 4. The y-axis is logarithmic to enhance the visibility of any deviations. Consequently the main scatter lies around zero and only a few engines show higher anomaly values. The numeric comparison between the three metrics is accomplished by their corresponding kurtosis value:

$$w = \frac{1}{N} \sum_{i}^{N} \left(\frac{\Delta_i - 1}{\sigma_i} \right)^4 \tag{7}$$

The mean value equals one for each metric as the values were already normalized to enable a direct comparison. The standard deviation σ_i corresponds to the scatter of anomaly values for the ensemble. A higher kurtosis value suggests a better separation of normal and abnormal machines, as the respective abnormalities deviate more from the normal distribution. Consequently, the residual metric performs slightly better than the euclidean metric. This is expected as it accounts for the typical scatter width. The kurtosis value of the reduced dc metric is approximately four times as high as the residual metric. This leads to the conclusion that the proposed rdc metric already outperforms classical approaches in its separation of anomalous engines by only comparing a single order-track to itself. As the rdc metric cannot only account for singular deviations of the mean value but consists of two-point differences, it is able to also spread its view onto more deviations.

E. Rdc metric in a hypermatrix

In this part the additional advantages of the rdc metric are highlighted. As it directly compares the relationship of two arbitrary engine orders to the typical relationship, its functionality can easily be expanded from a single engine

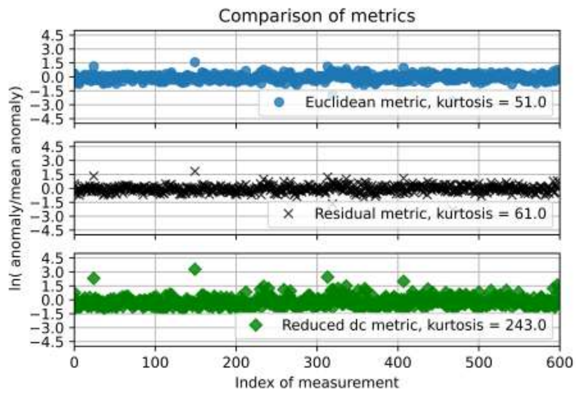


Fig. 4. Comparison of three metrics for detecting anomalies along a track of a fundamental gear meshing order. The rdc metric compares the same order with itself.

order to two different engine orders. Therefore it is possible to calculate rdcm-values for each order pair in the complete spectrogram (see figure 2). These can then be rearranged into a hypermatrix consisting of rdcm-values as colors. Their position describes which engine orders were used as input for each calculation. The rearrangement process is also depicted in figure 3. Therefore the whole spectrogram can be recalculated towards interconnected excitations and independent regions by the rdc metric.

V. Experimental examples

The evaluation of emerging patterns is conducted for two classes of faults. The first fault is a partially shorted winding in the electrically excited rotor. The second class revolves around faults that can be present at the C-bearing. All faults of the C-bearing strongly excite the fundamental order of the outer ring as a primary excitation. By asserting the secondary excitation mechanisms it is possible to directly name the fault at hand. The whole hypermatrix for a spectrogram

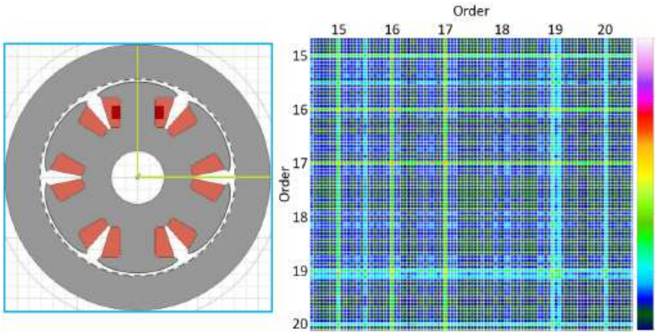


Fig. 5. Zoom into the resulting reduced dc-matrix for a partially shorted winding in the rotor. Approximately 10% of the windings were shorted. The sidebands of the 18th order emerge in equidistant patterns. At their cross points they show a strong interdependence. The 18th order itself remains nearly unaffected by the fault.

results in rather large images. Thus, for each fault a zoom is presented, highlighting the regions were the anomaly score is not near the minimum. A healthy machine does not exhibit any anomalous coincidences and is colored in the range of dark blue or black. As the figures of the faulty engines clearly show non-faulty regions between their emerging patterns, we do not add a separate, additional black image of a healthy engine.

A. *Partially shorted winding*

Electrically excited synchronous motors employ a rotor with copper windings. These windings are coated with an insulating layer to ensure a consistent number of windings for each pole. In our case the rotor consists of three pole pairs, consequently having six winding stacks. A schematic view of the rotor is given in figure 5. In order to simulate insufficient insulation between windings, inside a pole stack a manual short was introduced. It shorts approximately 10 % of the windings inside the stack. As the inductivity is directly proportional to the number of windings the reduction of 10% also is a good estimate in the case of no saturation. The resulting inconsistency in the flux leads to a estimated reduction in excitation proportional to the square of the magnetic field B^2. Therefore the relative amplitude of the 6th and 18th order is expected to be reduced by $10\,\log(0.9) = 0.45$ dB. This difference can not be distinguished from the normal scatter of the 600 units in the ensemble. As secondary excitations the sidebands due to a non-uniform magnetic field emerge and show a clearly visible pattern in the hypermatrix.

B. *Mechanical faults affiliated with the C-bearing*

Before going into detail of each fault, the fundamental orders are described. These are calculated in tab. II for the C- and A-bearing. All faults exhibit a strong anomaly score for the orders near the C-bearings outer ring. The classification towards different sources is therefore undertaken by the emerging patterns of the rdcm.

TABLE II
OVERVIEW OF THE ASSOCIATED ORDERS OF THE C AND A BEARINGS AND THE INPUT SHAFT

Order $\mathcal{O}$	C-bearing	A-bearing
$\mathcal{O}_{\mathrm{S-R}}$	1	
$\mathcal{O}_{\mathrm{C}}$	0.36	0.40
$\mathcal{O}_{\mathrm{bpo}}$	2.56	3.63
$\mathcal{O}_{\mathrm{bpi}}$	4.44	5.37
$\mathcal{O}_{\mathrm{b}}$	1.67	2.41
$\mathcal{O}_{\mathrm{em}}$	6	
$\mathcal{O}_{\mathrm{stator}}$	54	
$\mathcal{O}_{\mathrm{gm}}$	23	

1) Rough raceway: In the case of a rough raceway at the C-bearings outer ring, the fundamental harmonic is increased (see figure 6-*1)*). The balls inside the bearing excite the outer ring and its harmonics more strongly than in normal conditions. It is worth mentioning, that the rdc-matrix does explicitly show a normal behavior for the fundamental and first harmonic of the order corresponding to the C-bearings outer ring. Each raceway surface has a certain amount of roughness by design. Therefore, simply increasing the roughness does not change the behavior along the order track but introduces a broader excitation. Hence, the normal excitations are smeared towards neighbouring order lines. These result in anomalous tracks and add up to a higher anomaly score around the fundamental harmonic. There is no normal excitation mechanism for the higher harmonics as they are typically indistinguishable from normal noise. An increase above the noise level leads to newly anomalous tracks. Even in the case of only a small resonance band, which promotes the additional excitation above the noise floor a strong anomalous coincidence is found (visible at higher harmonics like $\mathcal{O} = 10.24$ or above).

2) Waviness corresponding to the number of balls inside the bearing: During the joining process and due to transportation issues it is possible that all balls inside a bearing press into the outer ring. In series production, this can happen if the outer ring blocks during its assembly, hence increasing the force needed to press the ring into the housing. This results in a waviness with a number of $Z = N$ balls. Therefore the resulting order directly coincides with the fundamental order of the outer ring. The emerging pattern is shown in figure 6-*2)*. As the force needed to press the balls into the outer ring can only be transmitted by the inner ring it is obvious that the same damage must exist at the inner ring to some extend. Therefore, the coinciding anomalies of the outer ring and inner ring match. Even if the amplitude of both excitations is not necessarily high the resulting order track changes in shape, consequently resulting in an increased rdc-value.

3) Misaligned shaft: As a final case of a faulty bearing situation inside the highly-integrated electric drive a misaligned shaft is chosen. The C-bearing is not aligned coaxially with the A-bearing resulting in a twist between

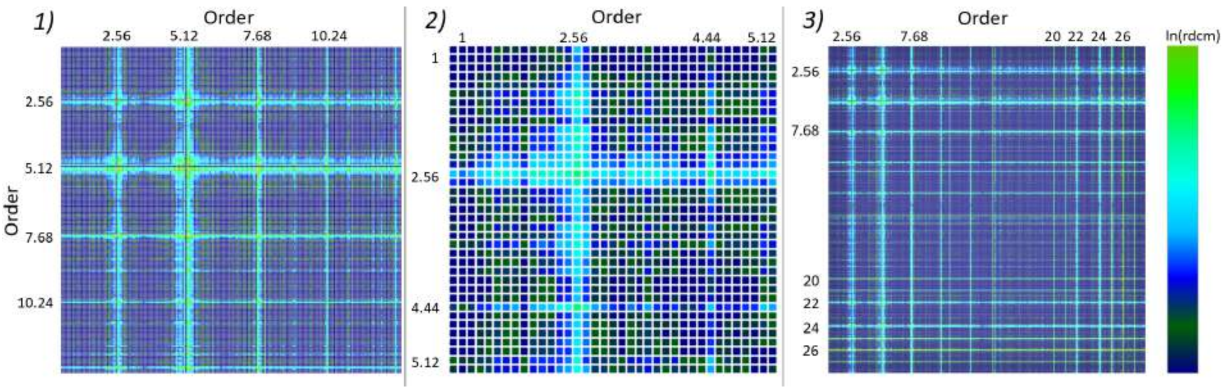


Fig. 6. Three cases that mainly excite the fundamental order of the c-bearing's outer ring. *1)* shows the corresponding rdc-matrix for a rough raceways. The rdc-matrix in *2)* exhibits a waviness at the outer ring. *3)* shows the effects of a misaligned shaft and its corresponding excitations.

the inner and outer ring. The connection between the input shaft and the A-bearing is implemented by a spline to the rotor shaft. As the rotor and its corresponding weight strongly stabilizes the rotor shaft and the spline can compensate for the misalignment, there is no twist present at the A-bearing. The complete rdc-matrix is displayed in figure 6-*3)*. The fundamental order of 2.56 and its harmonics is excited like in the case of a rough raceway (V-B1). Due to the misalignment, the gear meshing is modulated by the order $\mathcal{O}_\mathrm{S}$ resulting in equidistant sidebands of the 23th order. Especially at the crossing points of both ensembles of frequencies, an enhancement of the anomaly score is visible (upper right or lower left region in the figure).

VI. Discussion

The comparison of the ability to separate anomalies from the normal scatter is enhanced by the method of difference-correlations. Also, the rdc-metric approach for secondary excitations seems promising in unveiling emerging dependencies between different orders. As the basis functions of a Fourier transform are orthogonal by construction the resulting correlations are interpreted as purely physical phenomena. A limitation of this approach consist of a lack in experience for units with multiple faults. Multiple-fault scenarios lead to unknown pairs of correlating orders without physical reasoning. The interpretation of the rdc-matrices is therefore in need for a convincing validity check. This can either be achieved by a strong expertise in the field of fault diagnosis of rotating machinery. Another possibility for partially automating the classification is to divide the rdc-matrix into two sets of order pairs. The combination of all theoretically predicted order pairs forms the first set, all other pairs are sorted into the second set. The ratio between the sum of all anomaly scores in each set (divided by the number of pairs in each set) can then be asserted to. This covers the cases of theoretically predicted single faults or alarms an expert if the classification is not decisive enough. The latter can then be used to include the new fault into the first set. The second approach needs a strong confidence in knowing a huge proportion of possible fault combinations that are relevant for the unit to be tested.

VII. Conclusion

The classification of faulty components and bad assembly faces new challenges with increasing complexity and compactness of electric drives. The interplay of different excitation mechanisms results in primary and secondary symptoms. Common metrics are only able to capture primary excitations. Black-box approaches, such as methods from the domain of machine learning (e.g. neural networks, support vector machines etc.) are able to account for secondary symptoms but leave no option left for understanding their physical behavior. A white-box method is proposed to firstly detect anomalous order tracks and secondly set them in context of an accompanying order track. This allows to uncover secondary excitation mechanisms. These excitations build the foundation of classifying different faults, each sharing the same primary excitation. It is also possible to detect new emerging patterns in excited orders that were yet undiscovered due to individually weak signals. The reduced difference-correlation metric is compared to the common metrics (euclidean, residual), where it is shown that it is able to discover anomalies more clearly than its competitors. Afterwards, the reduced difference-correlation metric is employed to identify two classes of non-trivial faults of electrical and mechanical nature. The emerging patterns for a partially shorted winding are presented. For the class of mechanical faults, an ensemble of three different fault mechanisms is presented, each strongly exciting the bearings order of the outer ring. By including the context of secondary excitations visible in the reduced difference-correlation metric, one is capable of determining which fault was present at each unit. Future research will aim to automate the classification process based on the rdc metric. Assembling multiple faults into a single unit and comparing the emerging signatures to the case of single faults is in the scope of future work. The arising parallels and differences between different electric drive variants with a similar topology is also of high importance for the following considerations, as it may prove the generalizability of the approach.

REFERENCES

[1] G. Liu, J. Hong, and R. G. Parker, "Influence of simultaneous time-varying bearing and tooth mesh stiffness fluctuations on spur gear pair vibration," *Nonlinear Dynamics*, vol. 97, no. 2, pp. 1403–1424, 2019. [Online]. Available: https://doi.org/10.1007/s11071-019-05056-9

[2] P. Yang, Q. Yuan, C. Huang, Y. Zhou, H. Li, and Y. Zhou, "Analysis of the impacts of bearing on vibration characteristics of rotor," *Shock and Vibration*, vol. 2017, pp. 1–7, 12 2017.

[3] J.-W. Hur and T. A. Shifat, "Motor vibration analysis for the fault diagnosis in non- stationary operating conditions," vol. 12, pp. 151–160, 03 2020.

[4] A. Dineva, A. Mosavi, M. Gyimesi, I. Vajda, N. Nabipour, and T. Rabczuk, "Fault diagnosis of rotating electrical machines using multi-label classification," *Applied Sciences*, vol. 9, p. 5086, 11 2019.

[5] H. Yang, J. Mathew, and L. Ma, "Basis pursuit-based intelligent diagnosis of bearing faults," *Journal of Quality in Maintenance Engineering*, vol. 13, 06 2007.

[6] F. Immovilli, M. Lippi, and M. Cocconcelli, "Automated bearing fault detection via long short-term memory networks," 08 2019, pp. 452–458.

[7] A. Khlaief, K. Nguyen, K. Medjaher, A. Picot, P. Maussion, D. Tobon, B. Chauchat, and R. Cheron, "Feature engineering for ball bearing combined-fault detection and diagnostic," in *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)*, Aug 2019, pp. 384–390.

[8] V. Pandhare, J. Singh, and J. Lee, "Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features," 05 2019, pp. 320–326.

[9] A. Qu, B. VanHecke, D. He, and J. Yoon, "A study on comparing acoustic emission and vibration sensors for gearbox fault diagnostics," 05 2014.

[10] L. Shi, "Correlation coefficient of simplified neutrosophic sets for bearing fault diagnosis," *Shock and Vibration*, vol. 2016, pp. 1–11, 01 2016.

[11] S. Lu, Q. He, and J. Zhao, "Bearing fault diagnosis of a permanent magnet synchronous motor via a fast and online order analysis method in an embedded system," *Mechanical Systems and Signal Processing*, vol. 113, pp. 36 – 49, 2018, sI: IMETI-MechElectro. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327017301176

[12] C. Lessmeier, J. Kimotho, D. Zimmer, and W. Sextro, "Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification," 07 2016.

[13] J.-D. Wu, C.-W. Huang, and R. Huang, "An application of a recursive kalman filtering algorithm in rotating machinery fault diagnosis," *NDT & E International*, vol. 37, pp. 411–419, 07 2004.

[14] A. Yanjin and W. Hardt, "Automated noise analysis for fault detection," 06 2017.

[15] G. Meltzer and Y. Ivanov, "Fault detection in gear drives with non-stationary rotational speed - part ii: The time-quefrency approach," *Mechanical Systems and Signal Processing - MECH SYST SIGNAL PROCESS*, vol. 17, pp. 273–283, 03 2003.

[16] D. GmbH, *Anleitung der Discom-Messplattform*, 2020 (accessed December, 2020). [Online]. Available: https://manualzz.com/doc/4416084/tas-benutzerhandbuch

[17] P. McFadden and J. Smith, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique — a review," *Tribology International*, vol. 17, no. 1, pp. 3 – 10, 1984. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0301679X84900768

[18] M. VAISHYA and R. SINGH, "Analysis of periodically varying gear mesh systems with coulomb friction using floquet theory," *Journal of Sound and Vibration*, vol. 243, no. 3, pp. 525 – 545, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022460X00934199

[19] J. Liu, C. Yang, and Q. Lou, "Vibration analysis based feature extraction for bearing fault detection," *Applied Mechanics and Materials*, vol. 197, pp. 124–128, 09 2012.

[20] C. Adams, C. Schaal, J. Bös, and T. Melz, "Numerical investigation of the sound power and of the structural intensity of a permanent magnet synchronous machine," 08 2015.

[21] K. J. Meessen, J. J. H. Paulides, and E. A. Lomonova, "Force calculations in 3-d cylindrical structures using fourier analysis and the maxwell stress tensor," *IEEE Transactions on Magnetics*, vol. 49, no. 1, pp. 536–545, 2013.

VIII. BIOGRAPHIES

Jakob A. E. Bonart was born in Regensburg, Germany, in 1996. He received both his B.Sc. and M.Sc. in physics from Augsburg University, Bavaria in 2017/2019. During this time he partially worked as a researcher in the field of correlated quantum many-body systems at the chair TPIII, EKM, Augsburg. He is currently working on his PhD at BMW Group in cooperation with the Chemnitz University of Technology and Fraunhofer IWU, Dresden. His research activities focus on advanced statistical testing of electrical engines and correlations of structure-borne noise in electrical powertrains.

Joscha Eirich was born in Fürth, Germany, in 1990. He holds two degrees in Business Administration (B.Sc. and M.Sc.) as well as two in Information Systems (B.Sc. and M.Sc) from the University of Bamberg. During this time he partially worked as a researcher both at the chairs for statistics and information systems. He is currently working on his PhD at BMW Group in cooperation with the university of Bamberg. His research activities focus on the integration and quantification of domain knowledge into machine learning pipelines to improve the manufacturing of electrical engines and electrical energy storage systems.

Philipp Huemmer was born in Bamberg, Germany, in 1985. He received his M.Sc. in Mechanical Engineering and Applied Mechanics from University of Rhode Island and his Diploma from TU Braunschweig. He has spent his career in several areas of the testing world within the automobile industry. Since 2019, he is responsible for the development of testing procedures focusing involving structure-borne noise to detect errors in electrical powertrains.

Welf-Guntram Drossel was born in Freiberg, Germany, in 1967. He received his doctorate from the University of Technology Freiberg in 1988. After his PhD, he held several leadership positions at Fraunhofer-Institute for Machine Tools and Forming Technology, Chemnitz. In 2008 he became an Associate Professor at Chemnitz University and since 2014 he holds a full professorship. As of 2014 he also is the director of the Fraunhofer IWU.

6.10 P10: Kuenstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie

Künstliche Intelligenz: Anwendungen und Werkzeuge in die Automobilindustrie

Andre Luckow, Joscha Eirich, Kai Demtröder

Zusammenfassung

Künstliche Intelligenz (KI) kommt zunehmend im Alltag, in der Wissenschaft, aber auch in verschiedenen Bereichen der Automobilindustrie zum Einsatz. KI ermöglicht intelligenter Funktionen im Produkt und in Prozessen entlang der vollständigen automobilen Wertschöpfungskette. Zukünftig werden alle kritischen Anwendungen KI in unterschiedlicher Form nutzen.

Die Entwicklung und Einführung von KI-Anwendungen im industriellen Bereich ist aufgrund mehrerer Aspekte besonders herausfordernd: komplexe Datenflüsse und Abhängigkeiten und diverse Infrastrukturkomponenten bestehend aus Hardware und Software verteilt von der Cloud bis zur Edge müssen gemanagt werden. Eine der größten Herausforderungen ist die Erstellung praxistauglicher Machine Learning (ML) Modelle. Diese benötigen nicht nur große Mengen an Daten, sondern auch komplexe Aufbereitungs- und Labeling Prozesse.

Ziel dieses Kapitel ist es, einen Überblick von Anwendungsfällen der KI in der automobilen Wertschöpfungskette zu geben. Am Beispiel aus der Qualitätsanalyse in der Elektromotorenfertigung beschreiben wir als Mitarbeiter eines Automobilherstellers, wie die genannten Herausforderungen adressiert werden können. Wichtig ist hier insbesondere ein gesamthafter, interdisziplinärer Ansatz, welcher Technologien, Methoden, Tools, Prozesse und Kompetenzen vereint.

1 Einleitung

Die Automobilindustrie befindet sich in einem Wandel, getrieben durch technologische, gesellschaftliche und regulatorische Veränderungen. Neben der Transition zur Elektromobilität ist die Digitalisierung unserer Produkte eine Herausforderung. Heute erwarten Kunden modernste Technologien, sowie intelligente Funktionen und Dienste im Fahrzeug, wie sie nur der Einsatz von *Künstliche Intelligenz (KI)* ermöglicht werden. Gleichzeitig ist der Einsatz von KI in allen Prozessen entlang der automobilen Wertschöpfungskette wichtig, um Wettbewerbsvorteile zu erzielen, z. B. durch Prozessoptimierung eine Verbesserung der Kosten, Qualität und Effizienz [1].

Die Disziplin KI befasst sich mit der Erstellung von Systemen, die Aspekte von menschlichen, intelligenten Verhalten aufweisen. Unter dem Begriff werden i. d. R. eine Vielzahl von Technologien zusammengefasst, z. B. maschinelles Lernen, Optimierung, Simulation. Maschinelles Lernen ist aktuell das wichtigste Teilgebiet der KI und beschreibt statistische Lernverfahren, die aus Daten Muster extrahieren können und Prognosen zu erstellen.

Der Einsatz von KI in der Automobilindustrie unterliegt vielen Herausforderungen [2]. Die automobile Wertschöpfungskette umfasst ein komplexes Netzwerk von Lieferanten, Werken, Vertriebs- und Finanzdienstleistungsgesellschaften. Dementsprechend vielfältig und komplex sind existierende IT-Lösungen. Eine Erweiterung dieser um KI-Fähigkeiten ist oft schwierig. So kann die KI i. d. R. nur wenig von bereits vorhandenen Daten profitieren und müssen umfangreich, hochkomplexe und domain-spezifische Daten erheben. Zum Beispiel in der Qualitätssicherung sind oft hunderte bis tausend Beispieldatensätze für Defekte notwendig, um eine KI zu trainieren. Da Defekte aber äußerst selten auftreten, sind diese Daten nur sehr aufwendig zu beschaffen.

Ein Beispiel für die Prozesskomplexität in der Automobilfertigung sind die Umgebungsbedingungen, welche einen großen Einfluss auf das Produkt haben. So kann z. B. die Produktion elektrischer Antriebe durch externe Faktoren, wie Temperatur oder Luftdruck beeinflusst werden. Ein Fahrzeug, welches beispielsweise im Sommer lackiert wird, ist anderen Rahmenbedingungen ausgesetzt als im Winter. Diese unkontrollierbaren zusätzlichen Einflussgrößen können nur schwer mit Daten erfasst und von Modellen gelernt werden. Hieraus resultiert ein enormer zusätzlicher Modellierungsaufwand. Diese Herausforderung haben andere bekannte KI-Anwendungen nicht. Google's Go KI [3] kann z.B. die Art und Anzahl der Input für Modelle vollständig kontrollieren, alle Informationen des Systems vollständig erfassen und so eine entsprechend mächtige KI trainieren.

Des Weitern erfordert das Labeln dieser Daten und oftmals auch die Interpretation der KI-Ergebnisse ein hohes Maß an Domainexpertise. Die KI-Systeme müssen außerdem in existierende Infrastrukturen und damit Legacy Datenpipelines integriert werden. Dies ist nicht nur sehr aufwendig, sondern limitiert i. d. R. auch die Menge der gesammelten Daten, sowie die Möglichkeit Signale vom Benutzer (z. B. die manuelle Korrektur und Überschreiben von KI Ausgaben) zu sammeln.

Ziel dieses Kapitels ist es, einen Überblick (i) über Anwendungsfälle der KI zu geben, (ii) der dazu notwendigen Methoden, Technologien und Infrastruktur und (iii) Erfahrungen mit der Nutzung der Technologie. Dazu werden wir detailliert auf einen Anwendungsfall aus der Qualitätskontrolle im Bereich Elektromotorenfertigung eingehen und verschiedene Aspekte der KI, insbesondere die visuelle Analyse der Daten und Erzeugung hochqualitativer Labels und Modelle betrachten.

2 Hintergrund: Künstliche Intelligenz und Maschinelles Lernen

Der Begriff der *Künstlichen Intelligenz (KI)* wurde 1956 von John McCarthy und Marvin Minsky geprägt und umfasst verschiedene Technologien, welche das Ziel haben, Maschinen intelligentes Verhalten zu ermöglichen [4]. Eine dieser Technologien ist *maschinelles Lernen (ML)*. Im Folgenden fokussieren wir uns insbesondere auf Einsatzgebiete des maschinellen Lernens zum Verarbeiten von Daten, wie Sprache, Text, Sound, Bildern und Sensordaten. Prinzipiell ermöglicht maschinelles Lernen das Auffinden unterschiedlicher Arten von Mustern in den Daten, sowie die Gruppierung von ähnliche Daten. Diese Art des Lernens wird "unsupervised" genannt. So kann ein Algorithmus z. B. ähnliche Fehlermuster in unstrukturierten Sensordaten finden. Beim "supervised" Lernen wird das System mithilfe von konkreten Datenbeispielen trainiert und lernt so Prognosen und Klassifizierung auf automatisiert vorzunehmen.

Deep Learning (DL) ist eine Teildisziplin von ML und bezeichnet die Nutzung von komplexen neuronalen Netzwerken bestehend aus vielen unterschiedlichen Schichten [5]. Dabei haben sich abhängig von den verwendeten Datenarten und Aufgabe verschiedene Architekturen entwickelt, z.B. spezialisiert auf Computer Vision und Natural Language Processing Aufgaben. Fortschritte im Bereich Deep Learning werden vor allen getrieben durch die Verfügbarkeit von großen Datenmengen und Rechenkapazitäten, sowie den rapiden Fortschritt im Bereich DL Modellarchitekturen. Des Weiteren spielt Transfer Learning eine wichtige Rolle, welche die Wiederverwendung von vortrainierten Modellen ermöglicht und damit den Aufwand für den Einsatz von KI deutlich reduziert [6].

Computer Vision (CV) war eines der ersten Einsatzgebiete von Deep Learning. Convolutional Neural Networks (CNNs) ermöglichten durch die effiziente Komprimierung von Bilddaten unter Beibehalten wichtiger Struktur- und Kontextinformation in den Convolutional Schichten das effiziente Trainieren von neuronalen Netzen. CNNs werden daher für verschiedene Bildverarbeitungsaufgaben genutzt, z. B. Objekt-Klassifizierung, Erkennung, Lokalisierung und Segmentierung. Beispiele für CNN-Architekturen sind ResNet [7] und Inception [8]. Diese CNNs eignen sich für verschiedene Aufgaben im industriellen Bereich, z. B. die visuelle Inspektion in der Fertigung. Durch Transfer Learning kann der Bedarf an Trainingsdaten reduziert werden.

Im Bereich des Natural Language Processings (NLP) bringen Sprachmodelle in vielen Anwendungen große qualitative Vorteile, z. B. bei der Klassifizierung von Texten, maschinellen Übersetzungen, die Erstellung von Zusammenfassungen und dem Generieren von Texten. Die Basis hierfür sind sogenannte Transformer Sprachmodelle [9], wie GPT-3 [10] und BERT [11]. Diese werden auf existierenden Strukturen in Dokumenten und Texten trainiert. Durch das Ausblenden von Wörtern, dem sogenannten Masking, werden automatisch Labels generiert, welche dann als Prädiktionsziel für die Modelle verwendet werden. So können Modelle automatisch von der in Texten enthaltenen Struktur lernen und sich eine große Anzahl von Mustern aneignen. So können Sprachmodelle komplexe Repräsentationen der Daten lernen. Die Mächtigkeit von Sprachmodellen beruht insbesondere auch auf der Möglichkeit, diese auf den enormen Textmengen des Internets vorzutrainieren und dann auf verschiedenste Anwendungen zu adaptieren.

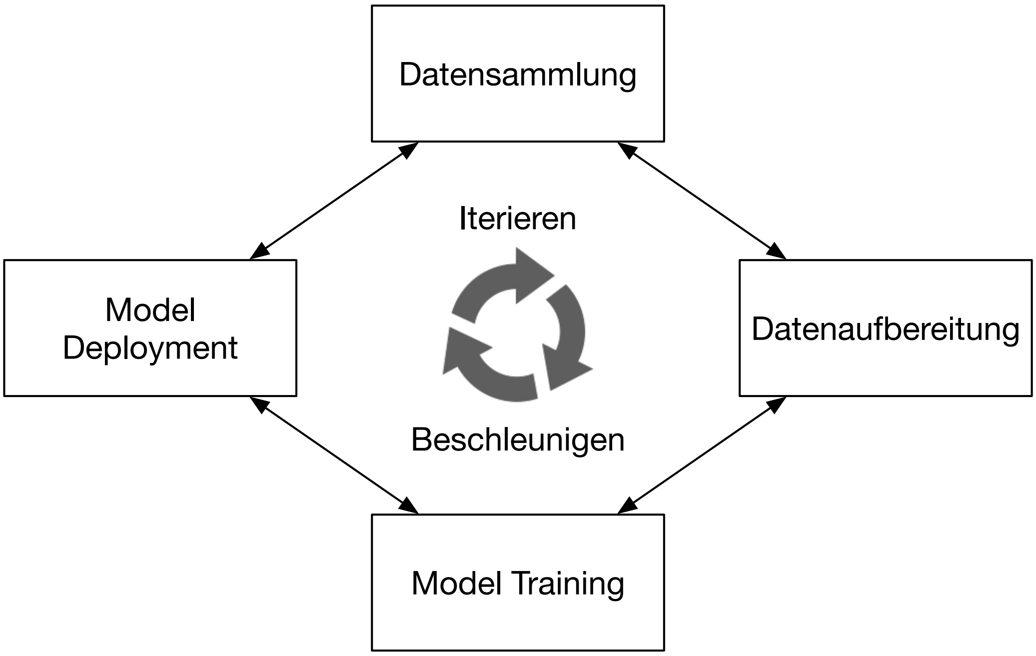


Abbildung 1: **KI Lifecycle:** MLOps Methoden und KI Plattform unterstützen den kompletten Lebenszyklus durch Abstraktion und Wiederverwendung.

2.1 KI Lifecycle

Abbildung 1 visualisiert den KI Lifecycle. Im Allgemeinen wird zwischen vier Schritten unterschieden, welche sehr eng miteinander verbunden sind.

(i) Datensammlung befasst sich mit der Erstellung von Daten. Hierbei kann man prinzipiell zwischen der Nutzung vorhandener Daten und der Beschaffung neuer aufgabenspezifischer Daten. Während Anwendungen mit explorativen Charakter oft mit vorhandenen Daten auskommen, ist es in den meisten Fällen und insbesondere zur Industrialisierung erforderlich neue Daten zu erheben. Dazu müssen robuste Datenpipelines implementiert und betrieben werden. Für Anwendungen mit Echtzeitanforderungen ist Streaming eine wichtige Fähigkeit, um solche Pipelines zu implementieren.

(ii) Datenaufbereitung beschreibt den Prozess der Transformation der Daten für ein spezifisches Modell, z. B. das Filtern, Aggregationen und Anonymisieren der Daten. Ein zentraler Teil der Datenaufbereitung ist das Feature Engineering, d. h. die Ableitung hochqualitativer Charakteristiken aus den Rohdaten, welche für das maschinelle Lernen genutzt werden können. Des Weiteren bedarf es, insbesondere für das supervised Lernen, so genannte Labels. Labels ordnen Daten eine bestimmte Klasse oder Kategorie zu und dienen beim Training des ML-Algorithmus als Ground Truth, welche es zu optimieren gilt. Die Erstellung solcher Labels erfordert gerade im industriellen Kontext i. d. R. ein hohes Maß an Domainwissen und Expertise. Oft werden hier spezifische Tools verwendet mit dem Ziel den Labeling-Prozess zu skalieren und zu industrialisieren.

Schritt (iii) befasst sich mit dem eigentlichen Trainieren des Models. Je nach verwendeten Modell sind hier spezifische Infrastrukturen, z. B. GPU Cluster, erforderlich. Im Schritt (iv) wird das Modell dann deployed und in die operative Anwendung integriert, z. B. durch einen Rollout an die Edge oder die Nutzung eines Cloud-Deployment Services. Edge Computing adressiert dabei die Notwendigkeit komplexe ML Modelle nahe der Datenquelle, d. h. Sensoren, Maschinen und anderer IoT Geräte, zu deployen, um die Performanz (z. B. die Latenz), die Datensicherheit und den Datenschutz zu erhöhen. Jedoch erhöht sich allerdings die Komplexität und damit der Management-Aufwand, da eine große Anzahl von Komponenten in einem hoch-verteilten System orchestriert werden müssen. Ein Beispiel für ein Edge-System sind KI Kamera-Systeme, die im Montageprozess zum Durchführen von visuellen Qualitätsinspektionen eingesetzt werden [12].

2.2 KI Infrastruktur

KI Applikationen sind komplexer und rechenintensiver als herkömmliche Applikationen und treiben die Anforderungen an IT-Infrastruktur, was Rechen- und Speicherkapazitäten, aber auch Software-Middleware betrifft. Die Rechenzeit, die für das Trainieren von KI Modellen genutzt wird, verdoppelt sich regelmäßig [13]. Dieser Trend ist primär getrieben durch die wachsende Größe von ML-Modellen, z. B. besitzt GPT-3 rund 175 Milliarden Parameter, die während des Trainings gehandhabt werden müssen. Das führt zu umfangreiche Anforderungen an eine Daten- und KI-Plattform, z.B. die Unterstützung diverse Datenquellen, Batch- und Streaming-Verarbeitung, Metadatenmanagement, Modellverwaltung und Deployment.

Mit der steigenden Bedeutung und Deployment von KI wurden die großen Herausforderungen der Industrialisierung sichtbar und die MLOps Disziplin gegründet [14]. MLOps hat das Ziel ganzheitlich alle Aspekte von KI Anwendungen zu adressieren und die Industrialisierung, Governance und Wartung von KI zu beschleunigen und professionalisieren. Damit verbunden ist die Notwendigkeit Software Engineering, Data Engineering, Machine Learning Modellierung und DevOps Practices zusammenzubringen. Neben methodischen Expertise in diesen Bereichen ist eine ausgereifte IT-Infrastruktur und Plattform erforderlich.

Für das Daten-Management, Analytics und Governance nutzen wir eine zentrale Daten-Plattform auf Basis von Cloud-Technologien von Amazon Web Services (AWS). Der *Cloud Data Hub (CDH)* [15, 16] ist eine zentrale Daten-Plattform, welche primär auf S3 als Storage-Technologie aufsetzt. Aufbauend auf diesem Fundament können verschiedenste AWS Dienste für den Daten-Ingest und Verarbeitung integriert werden, z. B. Athena für SQL und Glue für Spark-basierte Verarbeitung. Ein Datenportal bildet den Single-Point-of-Entry für End-User. Eine weitere zentrale Komponente ist der Datenkatalog, welcher in das Datenportal integriert ist und das Frontend zu einem zentralen Metadata-Datastore darstellt und es Nutzern ermöglicht Daten zu finden und abzufragen.

Die *KI-Plattform* erweitert den CDH, mit dem Ziel für die wachsende Anzahl an KI-Anwendungen eine wiederverwendbare und skalierende Infrastruktur zu schaffen, die es ermöglicht Machine Learning und Data Science Assets und Anwendungen, die diese nutzen, zu industrialisieren und zu betreiben. Ein wichtiger Aspekt ist z B. die Etablierung von Feedback Loops zwischen Datenströmen aus der Produktionsumgebung und der KI-Umgebung.

Die KI-Plattform bietet dafür verschiedene Templates, um aufbauend auf dem CDH wichtige MLOps Dienste, z. B. zum Model Training, Deployment und Betrieb, zu instanziieren und zu integrieren. Dabei liegt der Fokus auf KI-Eigenentwicklungen und der Nutzung möglichst gemanagter, höherer AWS Dienste, z. B. AWS Glue für Datentransformationen, SageMaker für Modellentwicklung und Deployment, sowie Lambda Funktionen zur Orchestrierung komplexer ML Worflows.

2.3 Visual Analytics: Domain-spezifische KI-Werkzeuge

Ein wichtiger Bestandteil von KI-Anwendungen sind Visualisierungswerkzeuge für verschiedene Zwecke: (i) Klassische *Business Intelligence (BI)* und (ii) Domain-spezifische *Visual Analytics (VA)* Tools.

BI Tools fokussieren sich dabei auf explorative Datenanalysen (z. B. zum Auffinden von Mustern und Generierung von Hypothesen), sowie die Erstellung von Dashboards. BI Werkzeuge unterstützen und automatisieren die Datensammlung, -Aufbereitung und Analyse mit dem Ziel, datengetriebene Unternehmensentscheidungen zu ermöglichen. Durch die Analyse dieser Informationen können im Nachgang entweder neue Informationen oder Wissen generiert werden. Traditionelle BI Anwendungen werden in der Regel für kleine und strukturierte Datenmengen angewandt, um Entscheidungen in Organisationen besser oder schneller treffen zu können. Es existieren eine Vielzahl von etablierten BI-Lösungen, welche in organisationalen Kontexten häufig verwendet werden. Zu den bekanntesten zählen Power BI, Tableau oder Qlik Sense.

Visual Analytics (VA) Werkzeuge fokussieren sich auf spezifische Domains und Anwendungsfälle und unterstützen besonders komplexe Datenstrukturen und Prozesse. Dazu werden diese VA-Tools oft eigens für diesen Zweck entwickelt und beinhalten User Interfaces, die eng auf einen bestimmten Kontext abgestimmt sind. Durch den Einsatz von Methoden der Statistik und KI bieten VA-Systeme eine Reihe von Möglichkeiten, den Benutzer zu unterstützen, z. B. durch das automatische Gruppieren von Daten, das Finden von ähnlichen Daten oder die Visualisierung von Mustern. VA Systeme können menschlichen Experten so besonders relevante Informationen vorschlagen [17]. Während VA-Systeme KI oft für verschiedene Benutzerfunktionen nutzen, bilden sie ebenfalls eine wichtige Grundlage für

Art	**Entwicklung**	**Produktion & Logistik**	**Kunde & Service**	**Indirekte Prozesse**
Typ I Data Science	Computer-Aided Engineering, z. B Computersimulationen und generatives Design, Lastenheftanalyse	Sensordatenanalyse, Produktions-KPI Analyse, Materialplannung	Kundenzufriedenheit und Loyalität, Preisoptimierung, Restwertoptimierung, Qualitätsanalysen	Process Mining, Vertragsanalyse, Lieferantenrisikobewertung, Maschinelle Übersetzung
Typ II Automatisierung	Intelligente Fahrzeugfunktionen, akustische Analyse von Fahrzeuggeräuschen	Visuelle Qualitätsprüfung, Predictive Maintance	Produktempfehlung, Intelligenter Persönlicher Assistent (IPA)	Conversational Systeme für interne Prozesse

Tabelle 1: KI in der Automobilen Wertschöpfungskette

die Erstellung von KI-Modellen, z. B. zur Erstellung hochqualitativer Trainingsdatensätze.

3 Anwendungen in der Automobilen Wertschöpfungskette

KI hält seit mehreren Jahren Einzug in die automobile Wertschöpfungskette [18, 19, 20, 21, 22] und generiert einen Mehrwert für Kunden, Produkte, Mitarbeiter und Prozesse [23]. KI wird durch den wachsenden Einsatz von Internet-of-Things Geräten und komplexen Sensordaten katalysiert.

Dabei geht der Trend zu preiswerten Sensoren und die Nutzung von KI-Methoden zur Aufbereitung der Daten. Diese Sensoren können optimal deployt und großflächig eingesetzt werden und ermöglichen somit eine umfangreiche Datensammlung. Dieses Vorgehen bedarf allerdings einer komplexeren KI- und Software-Infrastruktur für das Management von Datenflüssen, Modellen und Aktivitäten.

Prinzipiell unterscheiden wir zwei Arten der Nutzung: Data Science Anwendungen sind Typ I Anwendungen. Dabei wird KI für explorative Daten-Analysen genutzt. Hier sind oftmals Business-Fragen und das genaue Problem noch nicht genau bekannt. KI-basierte Verfahren werden hier genutzt, um Muster zu erkennen und darauf aufbauend Prognosen für Entscheidungen zu entwickeln. Der Mensch ist hier ein essenzieller Teil, denn nur er kann Mustern abstrahieren und zu kausalen Zusammenhängen verbinden und Wissen generieren. Typ I Anwendungen werden oft nur einmal problemspezifisch ausgeführt, z. B. eine Ursachenanalyse. Da hier Daten oft manuell gesammelt, aufbereitet und gelabelt werden, sind die Infrastrukturanforderungen nur gering.

Bei Typ II Anwendungen steht die Automatisierung und damit die Integration der KI in Geschäftsanwendungen und Prozesse im Vordergrund mit dem Ziel Entscheidungen zu unterstützen. So können wiederkehrende Entscheidungen auf Basis automatisiert werden und die Entscheidungsqualität und Konsistenz erhöht werden. Dazu sind neben dem eigentlichen ML, weiterführende Infrastrukturaspekte, wie Deployment, Integration, Betrieb und Datensicherheit essenziell.

Derzeit handelt es sich bei den meisten Anwendungen um Typ I Anwendungen, die zur Datenexploration und Analyse KI Methoden verwenden, um Entscheidungen zu unterstützen. Viele erfolgreiche Anwendungen werden kontinuierlich weiterentwickelt in Richtung KI-basierte Automatisierung (Typ II).

Tabelle 1 fasst eine kleine Auswahl von Anwendungsfällen zusammen, kategorisiert nach Anwendungstyp und Wertschöpfungssegment, d. h. Entwicklung, Produktion & Logistik, Kunde & Service und indirekten Bereich. KI hilft auf der einen Seite neue innovative Kundenfunktionen zu implementieren als auch auf der anderen Seite Prozesse durch Verbesserung in Qualität, Kosten und Zeit zu optimieren [24].

3.1 Entwicklung

Mit der zunehmenden Digitalisierung im und um das Fahrzeug nehmen die Datenmengen und damit Analysebedarfe im Entwicklungsbereich zu. Prinzipiell existieren drei Bereiche: (i) Computer-Aided Engineering (CAE), (ii) im Produkt selbst und (iii) zur Prozessoptimierung.

Computer-Aided Engineering ist ein instrumentaler Bestandteil bei der Entwicklung von modernen Fahrzeugen, z. B. zur Simulation der Fahrzeug-Aerodynamik, Unfällen und vielen anderen Aspekten. Fortschritte im Bereich Deep Learning ermöglichen z. B. die Nutzung von neuronalen Netzen für das

Approximieren von Simulationen und partiellen Differenzialgleichungen und die Auswertung von Simulationsdaten [25]. Damit kann sowohl die Güte von Computersimulationen, als auch die notwendigen Entwicklungs- und Rechenzeiten optimiert werden.

Im Produkt kommen heute intelligente Funktionen, wie der Intelligenter Persönlicher Assistent (IPA) und umfangreiche Sprachbedingungsmöglichkeiten zum Einsatz [23, 26]. Daneben spielt KI eine große Rolle bei modernen Fahrerassistenzsystemen, z. B. für das automatisierte Fahren [27].

Ein weiterer wichtiger Einsatzbereich von KI ist die Unterstützung der Fahrzeug-Entwicklungsprozesse, z. B. durch den Einsatz von NLP zur Analyse von Lastenheften und Spezifikationen im Anforderungsmanagement [28]. So existieren innerhalb der BMW Group mehr als 33.000 Lastenhefte mit über 30 Millionen Einzelanforderungen für die unterschiedlichen Baureihen und ihre Komponenten.

Dabei sind Anforderungen im Fahrzeugentwicklungsprozess der Startpunkt und die Basis für anschließende Funktionen, Wirk- und Lösungsprinzipien, bis hin zu Fertigungs- und Prüfkonzepten. Die KI-gestützte Prüfung der Spezifikationen ermöglicht es, die enormen Datenmengen schneller, einfacher und sorgfältiger zu verarbeiten, um so die Qualität von existierenden Dokumenten zu verbessern. So können, durch den Einsatz der KI, Anforderungen hinsichtlich Formalität, sprachlicher Qualität, Richtigkeit, Ähnlichkeit und Widerspruchsfreiheit geprüft werden. Da die Anforderungsdokumente auch Bestandteil von vertraglichen Vereinbarungen sind, ist es besonders wichtig, unscharfe Formulierungen, anhand von sogenannten "Weakwords", zu identifizieren und zu eliminieren. Die KI hilft dem Benutzer automatisiert Lastenhefte gezielt zu analysieren und somit kritische Passagen zu identifizieren und zu adressieren.

Zukünftige Entwicklungsprozesse lassen sich nur effizient gestalten, wenn die komplexen Wechselwirkungen zwischen unterschiedlichen Anforderungsformen und Zielen weiter automatisiert und manuelle Prüfungen hinfällig werden. Mit Hilfe KI-gestützter Methoden sollen Zusammenhänge zwischen den verschiedenen Ebenen und Funktions-, Lösungs- und Produktstrukturen der einzelnen Entwicklungsschritte hergestellt werden. Grundlage dafür ist, wie in den meisten Bereichen, eine durchgängige Datenbasis, die maschinell zu verarbeitenden Anforderungen und Beschreibungen enthält.

Dazu sollen vermehrt Sprachmodelle eingesetzt werden, die die Terminologie der Branche umfassen, um Anforderungen weiter zu systematisieren. Auf die Automobilbranche angepasste Sprachmodelle können hier genutzt werden, um ähnliche Anforderungen zu identifizieren. Neben Entitäten können zudem auch Zahlen und Einheiten im technischen Kontext identifiziert und ausgewertet werden. Ziel ist dabei die Repräsentation des Texts in Form von Entitäten und Beziehungen, um neben Anforderungen auch weitere Informationen mit einzubeziehen und zur Verfügung zu stellen, beispielsweise in einem Knowledge Graphen.

3.2 Produktion und Logistik

KI und Daten sind ein wichtiges Element von Industry 4.0 [29]. Unter dem Begriff Industrie 4.0 wird die Transformation klassischer Industrien, getrieben durch Digitalisierung und neue Technologien, wie KI, IoT, Robotik und Cloud Computing verstanden. Datenanalyseverfahren sind heute Standard um komplexe Produktionsabläufe zu verstehen und zu optimieren [30].

Ein Beispiel ist Predictive Maintenance von Fertigungsmaschinen und Robotern, mit dem Ziel Instandhaltungsprozesse und Abläufe zu optimieren. Dazu werden KI-basierte Fehlervorhersagen verwendet, um abnormales Verhalten zu identifizieren und durch frühzeitige Wartung potenzielle Maschinenausfallzeiten zu vermeiden [31]. Eine Herausforderung ist dabei die Verfügbarkeit und Interpretation der Sensordaten dieser Maschinen, welche oft noch nicht für den Einsatz von KI ausgelegt worden sind.

Ein weiterer wichtiger Bereich für KI ist die Qualitätskontrolle, z. B. die automatisierte Analyse von Sensordaten von Testgeräten oder die Nutzung KI-basierter Bilderkennungsverfahren [19]. Diese Bilderkennungsverfahren ermöglichen z. B. die automatisierte Erkennung und Prüfung von bestimmten Fahrzeugmerkmalen, wie das Vorhandensein von Warnetiketten, den korrekten Teilen (z. B. Felgen, Befestigungsstopfen) und den korrekten Typschildern. Dazu werden immer mehr Kameras in verschiedene Bereiche der Produktion eingesetzt. Dabei kommen verschiedene KI-Verfahren zum Einsatz, z. B. zur Objekterkennung, Segmentierung, Klassifizierung, Anomalie-Erkennung und OCR.

3.3 Kunde & Service

KI spielt eine zunehmende wichtige Rolle im Kundenbereich. Neben den genannten Kundenfunktionen, kann KI Aspekte wie Benutzerempfehlung (z.B. Ausstattung), Risikobewertung bei Kredit- und

Leasingverträgen, und im Aftersales Bereich zur Analyse von Qualitätsthemen unterstützen.

Ein Ziel dieser Analysen ist es, potenzielle Probleme frühzeitig zu identifizieren und die Kundenzufriedenheit [32] zu verbessern. Dies ist wichtig, um langfristig erfolgreich im Markt zu sein und die Loyalität von Kunden zu fördern. Die Kundenzufriedenheit wird von vielfältigen Faktoren bestimmt, z.B. der Qualität des Produktes und Services, aber auch Aspekten wie Kulanz. Dazu müssen Daten aus verschiedensten Bereichen des Unternehmens integriert und analysiert werden.

3.4 Indirekter Bereich

Conversational KI findet nicht nur im Produkt ein wichtiges Einsatzgebiet, sondern unterstützen auch immer mehr Prozesse und bietet eine einheitliche Benutzerschnittstelle zu Funktionen wie Mitarbeiterinformationen, Meeting- und Raumplannung und FAQs [28].

In globalen Unternehmen spielen aufgrund der Vielfältigkeit der Sprachen Übersetzungen eine zentrale Rolle, um sowohl Kunden, Lieferanten, Partnern, Händlern als auch Mitarbeitern weltweit Informationen zur Verfügung zu stellen. Täglich erreicht das Unternehmen eine große Menge an mehrsprachigen Texten aus externen Quellen. Dabei ist die menschliche Übersetzung aller Daten aufgrund des Volumens, der Vielfalt und auch der Kosten oft keine Option mehr.

Maschinelle Übersetzung, welche bereits heute auch im privaten Leben vielfach und einfach eingesetzt werden kann, scheint hier die Lösung, um Übersetzungszeiten mit der Reduktion der manuellen Arbeit zu verringern und so dem Unternehmen zu helfen, schneller und effizienter zu arbeiten. Jedoch eignen sich frei zugängliche Off-the-Shelf Produkte meist nur bedingt. Zum einen entsprechen sie nicht immer dem geforderten Sicherheitsniveau, welches für manche sensiblen Daten aus dem Produktionsbereich erforderlich ist. Zum anderen haben sie oftmals eine schlechte Performance für die teilweise sehr spezielle Terminologie und technische Ausdrucksweisen im Automobilbereich. Im Rahmen des "Customized Machine Translation" Projektes entwickeln wir ein Übersetzungssystem, welches auf die sehr spezifischen Bedürfnisse des Fachjargons in der Automobilindustrie eingeht und Übersetzungen für andere Anwendungen anbietet.

Dazu wurden eigene Modelle mit der spezifischen Terminologie und eigenen Sprachkorpera trainiert, um manuelle Übersetzung oder die nachträgliche Bearbeitung durch einen "Human-in-the-loop" im sogenannten "post-editing" auf ein Minimum zu reduzieren. Dabei findet die maschinelle Übersetzung Anwendung in zahlreichen Use Cases in den unterschiedlichsten Fachbereichen. So können bereits heute Änderungsanforderungen für alle Produktionswerke weltweit, täglich automatisiert übersetzt und verteilt werden. Zudem findet sich eine große Anwendung im Aftersales-Bereich, um die Kommunikation mit Kunden und Händlern weltweit zu vereinfachen und zu beschleunigen. Immer mehr Use Cases sind dabei auf die maschinelle Übersetzung angewiesen, wie beispielsweise die Auswertung von Social Media Daten aus unterschiedlichsten Quellen. Grundlage für eine gute, automatisierte Übersetzung ist dabei nicht unbedingt der Algorithmus, welcher auf sogenannten Encoder und Decoder Modellen basiert, sondern die Qualität der zugrunde liegenden Trainingsdaten.

3.5 Diskussion

Automobile KI Anwendungen sind sehr vielfältig und komplex: (i) die architekturelle Komplexität von ML Modellen ist sehr hoch, sowohl was das Trainieren als auch das Deployment angeht, (ii) der Notwendigkeit komplexer und hochqualitativer Datensätze und Trainingsdaten (inkl. Labels) und (iii) der hohe Aufwand für das Deployment von KI-Anwendung und die dafür notwendige Integration von Datenflüssen, Trainings-, Inference-Umgebungen, sowie Prozess-IT-Systemen.

Sowohl Sprach- als auch andere Deep Learning Modelle besitzen eine große Anzahl von trainierbaren Parametern und erfordern enorme Rechenressourcen zum Trainieren, als auch für die Inferenz, d. h. die Anwendung des Modells auf eine spezifische Aufgabe. Die eigentliche ML Komponente repräsentiert in der Regel nur einen kleinen Teil der eigentlichen Applikation. Insbesondere Daten-Abhängigkeiten stellen eine große Herausforderung für das Deployment von KI-Anwendungen dar [33]. Hier sind die klassischen Herausforderungen in den 4V's definiert: Volumen, Velocity, Variety, und Veracity. Eine der großen Herausforderung der KI ist die semantische Interpretierbarkeit der Daten und der Verfügbarkeit von Labels, welche Daten auf standardisierte Konzepte zuordnen.

Für die Nutzung von KI sind i. d. R. wichtige Daten aus Legacy Systemen erforderlich, z. B. Prozessinformationen. Des Weiteren müssen Ergebnisse in Prozesssystemen weiterverarbeitet werden.

Gleichzeitig werden KI Systeme oft im Kontext des Internet-of-Things eingesetzt, um die dort generierten Daten zu verarbeiten. Daraus resultiert sowohl eine architekturelle als auch operative Komplexität, die erforderlich ist, um Prozess-, IoT- und KI-Systeme zu integrieren. Hinzu kommt, dass es bis heute kein einheitlicher Standard existiert, um besonders IoT Daten in einheitlichen Formaten Nutzern zur Verfügung zu stellen.

Die weitere Industrialisierung von KI erfordert das Neudenken von Prozessen, Mensch-Maschine Interaktionen (z. B. was das Labeln von Daten betrifft) und eine skalierbare KI Infrastruktur. Dabei spielt insbesondere die Notwendigkeit hochqualitative Daten zu sammeln, aufzubereiten und labeln eine wichtige Rolle. Im Folgenden stellen wir dazu ein spezifisches Visual Analytics Tool vor, welches zur Analyse und Aufbereitung von Qualitätsdaten aus der Elektromotorenfertigung entwickelt wurde.

4 Zusammenfassung

KI und Machine Learning Methoden sind in der Wirtschaft und Industrie angekommen. Die Nutzung dieser zur Analyse von wachsenden Datenmengen ermöglicht Unternehmen in der Automobilindustrie einen Wettbewerbsvorteil durch bessere Entscheidungen (Typ I Data Science) und zunehmend die Automatisierung von Geschäftsprozessen und Entscheidungen (Typ II Automatisierung).

Die Verfügbarkeit von aufgaben-spezifischen, hoch-qualitativen Daten ist essenziell, um akkurate und robuste Modelle zu erstellen. Dabei ist die Akquise solcher Daten, insbesondere in speziellen Industrieumgebungen, herausfordernd und erfordert eine komplexe IT-Infrastruktur und vielfältiges Expertenwissen. Wie am Beispiel Körperschallprüfung gesehen, sind Visual Analytics Systeme, wie IRVINE, ein essenzielles Bindeglied zwischen Domain Expertise und KI Methodik. Die Verbindung von einer unsupervised ML mit Expertenfeedback ermöglicht die Erkennung und Verifizierung neuer Muster. Damit werden hoch-qualitative Labels geschaffen und Unternehmenswissen in KI Modelle überführt. Langfristig können diese Modelle nicht nur hoch-spezialisierte Domainexperten, die Maschine und Prozess kennen, unterstützen, sondern in KI Prädiktionsverfahren überführt werden. Dadurch kann Expertenwissen skaliert und breit im Feld genutzt werden, um somit frühzeitig potenzielle Probleme zu erkennen und Mitarbeiter in Entscheidungen zu unterstützen. Dadurch wird die Entscheidungsqualität und Konsistenz verbessert.

Machine Learning ist ein mächtiges Werkzeug, jedoch nicht die Lösung für alle Probleme. Oftmals können einfache Heuristiken schon einen Vorteil bieten. Die Komplexität von KI-Systemen und deren Realisierung sollte nicht unterschätzt werden. MLOps bietet hier das notwendige Rahmenwerk, um Domainexpertise, Data Engineering, Data Science, Software Engineering zu vereinen und KI ganzheitlich anzugehen.

Zukünftig planen wir eine Vielzahl von KI-Anwendungen zu industrialisieren und KI breiter im Unternehmen auszurollen. Gleichzeitig ist der verantwortungsvolle Einsatz von KI sicherzustellen und wichtig Grundsätze, die den Einsatz von KI regeln zu etablieren. Dazu zählt z. B. die Sicherstellung der Privatsphäre und des Datenschutzes (z. B. durch Verwendung von KI Anonymisierungsverfahren [34]), der Vorrang menschlicher Handels und maximale Sicherheit und Transparenz [35].

Danksagungen

Literatur

[1] Andreas Cornet, Matthias Kässer, Thibaut Müller, and Andreas Tschiesner. Künstliche Intelligenz revolutioniert Autosektor. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-road-to-artificial-intelligence-in-mobility-smart-moves-required/de-DE, 2017.

[2] Andrew Ng. Ai doesn't have to be too complicated or expensive for your business. *Harvard Business Review,* *https://hbr.org/2021/07/ai-doesnt-have-to-be-too-complicated-or-expensive-for-your-business*, 2021.

[3] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Made-

leine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[4] Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon, and John McCarthy. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. 27(4):12, 2006.

[5] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[6] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Kohd, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models, 2021.

[7] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[8] Christian Szegedy, Sergey Ioffe, and Vincent Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. *CoRR*, abs/1602.07261, 2016.

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[10] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[12] Aishwarya Srivastava, Dung Nguyen, Siddhant Aggarwal, Andre Luckow, Edward Duffy, Ken Kennedy, Marcin Ziolkowski, and Amy Apon. Performance and memory trade-offs of deep learning object detection in fast streaming high-definition images. In *2018 IEEE International Conference on Big Data (Big Data)*, Dec 2018.

[13] Dario Amodei and Danny Hernandez. Ai and compute. https://openai.com/blog/ai-and-compute/, 2018.

[14] Cristiano Breuel. Ml ops: Machine learning as an engineering discipline. Towards Data Science, https://towardsdatascience.com/ml-ops-machine-learning-as-an-engineering-discipline-b86ca4874a3f, 2020.

[15] BMW Group. Cloud data hub: Organize all of bmw group's data and make it universally actionable. https://data.bmwgroup.com/, 2021.

[16] Josef Viehhauser, Simon Kern, and Junjie Tang. Creating a data-driven, cloud-native ecosystem at the bmw group. ReInvent Presentation available at: https://d1.awsstatic.com/events/reinvent/2019/Creating_a_data-driven,_cloud-native_ecosystem_at_BMW_Group_AUT306.pdf, 2019.

[17] James J. Thomas and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[18] Andre Luckow, Ken Kennedy, Fabian Manhardt, Emil Djerekarov, Bennie Vorster, and Amy Apon. Automotive big data: Applications, workloads and infrastructures. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 1201–1210, 2015.

[19] Andre Luckow, Ken Kennedy, Marcin Ziolkowski, Emil Djerekarov, Matthew Cook, Edward Duffy, Michael Schleiss, Bennie Vorster, Edwin Weill, Ankit Kulshrestha, and Melissa C Smith. Artificial intelligence and deep learning applications for automotive manufacturing. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 3144–3152, 2018.

[20] Joscha Eirich, Dominik Jäckle, Tobias Schreck, Jakob Bonart, Oliver Posegga, and Kai Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *VDS - Symposium on Visualization in Data Science at IEEE VIS*, 11 2020.

[21] Joscha Eirich, Dominik Jäckle, Stefan Werrlich, and Tobias Schreck. Visual analytics in organization knowledge creation: A case study. In *European Conference on Information Systems*, 04 2021.

[22] Joscha Eirich, Jakob Bonart, Dominik Jäckle, Michael Sedlmair, Ute Schmid, Kai Fischbach, Tobias Schreck, and Jürgen Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. *IEEE Transactions on Visualization and Computer Graphics*, 07 2021.

[23] BMW Group. Künstliche intelligenz: Schlüsseltechnologie der digitalen transformation. https://www.bmwgroup.com/de/innovation/technologie-und-mobilitaet/artificial-intelligence.html, 2020.

[24] Peter Buxmann and Holger Schmidt, editors. *Künstliche Intelligenz*. Springer Gabler, 2018.

[25] Karen Hao. Ai has cracked a key mathematical puzzle for understanding our world. https://www.technologyreview.com/2020/10/30/1011435/ai-fourier-neural-network-cracks-navier-stokes-and-partial-differential-equations/, 2020.

[26] Alvin Chin, Jilei Tian, and Johann P. Prenninger. Toward contextual and personalized interior experience in a vehicle: Predictive preconditioning. In *92nd IEEE Vehicular Technology Conference, VTC Fall 2020, Victoria, BC, Canada, November 18 - December 16, 2020*, pages 1–5. IEEE, 2020.

[27] BMW Group. Autonomes fahren: Intelligenz, die uns bewegt. https://www.bmwgroup.com/de/innovation/technologie-und-mobilitaet/autonomes-fahren.html, 2018.

[28] Werner Beutnagel. Anwendung in der Business-IT: Natural Language Processing wird bei BMW zum stetigen Helfer. Automotive IT, https://www.automotiveit.eu/strategy/natural-language-processing-wird-bei-bmw-zum-stetigen-helfer-108.html, 2020.

[29] Human Ramezani and Andre Luckow. Big data, small data, and getting products right first time. In Mohammad Dastbaz and Peter Cochrane, Industry 4.0 and Engineering for a Sustainable Future: https://doi.org/10.1007/978-3-030-12953-8_6, 2019.

[30] Sandeep Jeereddy, Ken Kennedy, Eddie Duffy, Annie Walker, and Bennie Vorster. Machine learning use cases for smart manufacturing kpis. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4375–4380, 2019.

[31] BMW Group. Predictive maintenance: When a machine knows in advance that repairs are needed. https://www.press.bmwgroup.com/global/article/detail/T0338859EN/predictive-maintenance:-when-a-machine-knows-in-advance-that-repairs-are-needed, 2021.

[32] Stefan Meinzer, Ulf Jensen, Alexander Thamm, Joachim Hornegger, and Björn Eskofier. Can machine learning techniques predict customer dissatisfaction? A feasibility study for the automotive industry. *Artificial Intelligence Research*, 6(1):80–90, 2017.

[33] D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[34] Jimmy Tekli, Bechara al Bouna, Raphaël Couturier, Gilbert Tekli, Zeinab al Zein, and Marc Kamradt. A framework for evaluating image obfuscation under deep learning-assisted privacy attacks. In *17th International Conference on Privacy, Security and Trust, PST 2019, Fredericton, NB, Canada, August 26-28, 2019*, pages 1–10. IEEE, 2019.

[35] BMW Group. Bmw group code of ethics for artificial intelligence. https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/downloads/ENG_PR_CodeOfEthicsForAI_Short.pdf, 2020.

Full reference: A. Luckow, J. Eirich, and Demtroeder. K. Kuenstliche Intelligenz: Anwendungen und Werkzeuge in der Automobilindustrie. In *Wie funktionier Data Science?*, pp. 1–11. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

6.11 P11: Visual Analytics zur Optimierung von Pruefprozessen in der Serienfertigung elektrischer Motoren

Full reference: J. Eirich, A. Luckow, and Demtroeder. K. Visual Analytics zur Optimierung von Pruefprozessen in der Serienfertigung elektrischer Motoren. In *Wie funktionier Data Science?*, pp. 1–7. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

Visual Analytics zur Optimierung von Prüfprozessen in der Serienfertigung elektrischer Motoren

Joscha Eirich, Andre Luckow, Kai Demtröder

Zusammenfassung

Fortschritte in Technologien wie dem Internet der Dinge (IoT) und Visual Analytics (VA) führen zu neuen Möglichkeiten in der intelligenten Fertigung. Diese Technologien bieten enorme Potenziale zur Verbesserung von Produktionsprozessen, wie zum Beispiel eine gesteigerte Qualität der produzierten Teile oder eine erhöhte Taktzeit. Die Herausforderung bleibt jedoch, hochdimensionale und komplexe IoT-Daten menschlichen Experten für Analysen leicht zugänglich zu machen. In diesem Kapitel stellen wir die Anwendung einen Visual Analytics Systems auf die Körperschallprüfung elektrischer Motoren bei der BMW Group vor. Ziel hierbei ist es zu zeigen, wie anhand der Informationsvisualisierung mittels Visual Analytics Fachexperten in ihren oftmals komplexen Analysen unterstützt werden können, um somit die Qualität der Serienfertigung zu steigern.

1 Einleitung

Während der Fertigungsprozess werden Daten auf verschiedenen Ebenen und mit unterschiedlichen Granularitäten generiert, z. B. Vibrations-, Temperatur- und Kamera-basierte Sensoren auf Maschinen-Ebene und Transaktions-, Prozess- und Unternehmensdaten auf Prozessebene. Im Folgenden fokussieren wir uns auf die Analyse von Qualitätsmessdaten, welche im Rahmen der Qualitätsprüfung für Elektromotoren mit einem Körperschallmessgerät erhoben werden. Die so erhobenen Daten sind hochkomplex und bedürfen eines hohen fachlichen Maßes an Expertise bei der Auswertung. Bei der Analyse dieser Daten bestehen jedoch die folgenden zwei Herausforderungen:

- In der Serienfertigung werden oft mehrere tausend Motoren pro Tag in einem Werk produziert. Eine detaillierte Analyse der Daten jedes einzelnen Motors durch einen Experten ist somit nicht möglich.
- Da nur wenige Motoren im Detail von Experten analysiert werden können, ist es umso wichtiger, dass das aus Analysen resultierende Wissen nachhaltig gespeichert und anderen Mitgliedern der Organisation zur Verfügung gestellt wird.

Um eine Vielzahl an komplexen Daten von Motoren einfach analysierbar zu machen und das resultierende Wissen zu speichern, setzt BMW zunehmend auf neue Verfahren der Informationsvisualisierung wie Visual Analytics (VA). VA ist ein Zweig der Informationsvisualisierung, der sich auf die interaktive Visualisierung von Informationen konzentriert. VA umfasst hierbei verschiedene Aspekte analytischer Aufgaben wie etwa die Vorverarbeitung großer unstrukturierter Datenmengen, das Erstellen von KI Modellen sowie automatisierte Empfehlungen von besonders interessanten Datenpunkten für eine detaillierte Analyse [1]. Ein beispielhafter Prozess, wie VA Experten in der Analyse von Produktionsdaten unterstützen kann, ist in Abbildung 1 dargestellt. Hier werden Sensor-Daten direkt in der Serienfertigung aufgezeichnet und an ein VA System geschickt. Experten nutzen die visuellen Abstraktionen der Produktionsdaten im VA System um einerseits selber die Daten besser zu verstehen und anderseits produzierte Bauteile anhand von Labels klassifizieren. So kann ein Bauteil etwa das Label *in Ordnung* oder *nicht in Ordnung* haben. Erstellte Labels können anschließend genutzt werden, um KI Modelle zu trainieren, welche nun vollautomatisch Sensor-Daten analysieren und eine Aussage bezüglich ihrer Qualität treffen können. Hierbei gilt: Je mehr Labels erstellt wurden, desto besser können KI Modelle Vorhersagen treffen. Somit kann mittels VA eine Human-in-the-loop Schnittstelle zwischen Mensch und Maschine erstellt werden.

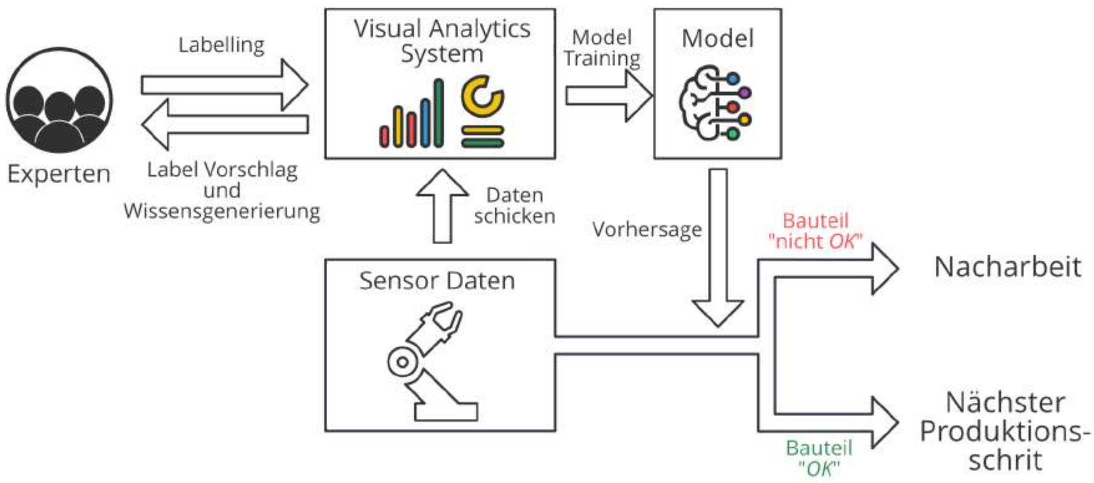


Abbildung 1: **Visual Analytics (VA) Körperschall Prozess:** Das Visual Analytics (VA) System bereitet hoch-komplexe Sensordaten auf. Die vorgeschlagenen Muster dienen als Basis für die Akquise von Labels, welche für die Erstellung von ML Modellen dienen.

Um die Analyse komplexer Produktionsdaten zu erleichtern, wurde innerhalb BMWs das Visual Analytics (VA) System IRVINE entwickelt [2]. Das Ziel von IRVINE ist es, Fachexperten in der Analyse von Körperschalldaten zu unterstützen, um Fehler in Motoren besser detektieren und deren Ursache verstehen zu können. Im Rahmen dieser Analyse nutzt IRVINE verschiedene KI Methoden und schlägt Experten besonders relevante Motoren für eine Detailanalyse vor. Experten können dem Motor dann ein Label zuordnen und die Ursache des Fehlers (im folgenden "Annotation") in den Messdaten markieren. Nach jeder Analyse wird das so generierte Wissen in Form von Labels und Annotation externalisiert und in einer Datenbank gespeichert. Diese hochqualitativen Trainingsdatensätze sind besonders für spätere Modellentwicklungen relevant.

Dieses Kapitel ist wie folgt gegliedert: In Abschnitt 2 beschreiben wir das Messverfahren "Körperschall". Im Anschluss (Abschnitt 3) gehen wir auf die Abstraktion der Daten ein, welche im Körperschall aufgezeichnet werden. In Abschnitt 4, erläutern wir, wie KI genutzt wird, um Experten zu unterstützen. Das fertige IRVINE System wird in Abschnitt 5 beschrieben.

2 Körperschallprüfung

Eine Körperschallprüfung beschreibt den Schall, welcher sich in einem Motor ausbreitet. In dieser Art Prüfung wird die Motordrehzahl unter kontrollierten Bedingungen gesteuert, um das Verhalten von Motoren und deren Bauteilen zu analysieren. Die Schallmessung ermöglicht die Beurteilung der akustischen Eigenschaften von Bauteilen und ihres technischen Zustands. Ein Beispiel ist das künstliche Drehen des Motors, um dessen Beschleunigungsverhalten von 0 auf 100 km/h zu überprüfen. So kann festgestellt werden, ob sich ein Motor wie erwartet verhält oder teilweise zu laut ist. Eine erhöhte Lautstärke ist hier ein Indiz für einen Schaden am Motor. Das Ergebnis solcher akustischer Messungen ist eine dreidimensionale Datenstruktur bestehend aus *Lautstärke*, gemessen über die *Drehzahl* und *Ordnungen*.

Eine beispielhafte Körperschallmessung ist in Abbildung 2 als sogenanntes Spektrogramm dargestellt. Auf der Y-Achse wird die Drehzahl angezeigt, welche die Beschleunigung des Motors bis zur Höchstgeschwindigkeit beschreibt. Wenn ein Motor also von 0 auf 100 km/h beschleunigt, erhöht sich seine Drehzahl entsprechend von bspw. 2000 Umdrehungen pro Minute bei 20 km/h auf 8000 Umdre-

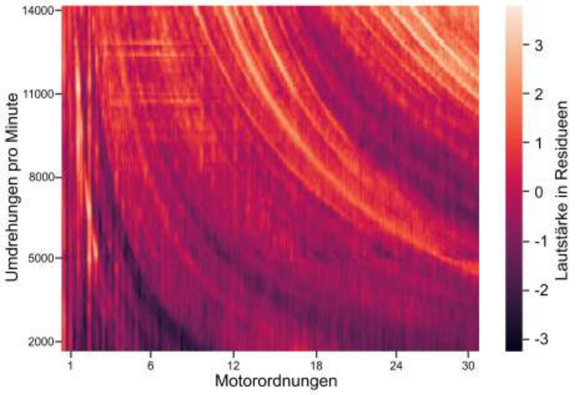


Abbildung 2: **Dreidimensionales Spektrogramm einer Körperschallmessung:** Auf der Y-Achse ist die Drehzahl angetragen. Die X-Achse zeigt die Motorordnungen und die Farbe den Residualwert, also den Schweregrad der aufgetretenen Anomalie.

hungen pro Minute bei 100 km/h. Im Körperschall sind manche Fehler in bestimmten Drehzahlbereichen besonders gut zu erkennen. So kann ein Fehler an der Getriebewelle etwa besser in höheren Drehzahlen erkannt werden, da sich dort die Getriebewellen schneller drehen und somit beschädigte Wellen lauter zu hören sind als im niedrigen Drehzahlbereich.

Auf der x-Achse sind die Ordnungen dargestellt, die das Verhältnis zwischen einer gemessenen Frequenz und der Drehzahl des Motors während der Messung darstellen und somit beschreiben, wie oft pro Umdrehung eine Anregung auftritt. Ordnungen sind von Vorteil, da sie eine detaillierte Analyse von Teilkomponenten ermöglichen. Durch die Analyse der Geometrie des rotierenden Motors sind Ingenieure in der Lage, die vermessenen Teilkomponenten eines Motors bis auf eine sehr detaillierte Ebene zu analysieren (z. B. die Analyse des 24ten Zahns eines Zahnrads). Innerhalb eines Spektrogramms beschreibt die X-Achse, also die Komponente eines Motors, die analysiert wird. So kann bspw. in der Spalte der 1ten Motorordnung ein Rückschluss auf die Rotorwelle getroffen werden und in der 6ten das A-Lager eines Motors analysiert werden.

Die Körperschalldaten, die in Dezibel aufgezeichnet werden, werden außerdem in Residuen umgewandelt. Der Vorteil von Residuen ist, dass sie keine Absolutwerte, sondern die Abweichung im Verhältnis zu einer Grundgesamtheit an Motoren darstellen. So können vor allem statistische Ausreißer besser erkannt und analysiert werden. Die Lautstärke in Residuen wird auf einer Farbskala dargestellt, wobei hellere Farben besonders laute Ordnungen darstellen. Regionen mit lauten Ordnungen sind dabei ein guter Indikator für unterschiedliche Fehlerarten. Die Residuen (die Farbe) beschreiben den Schweregrad eines Fehlers. So sind helle Farben ein Indikator für einen Fehler.

Ähnlich wie bei anderen verwandten Datenstrukturen, z. B. aus der Sonografie bzw. des Ultraschalls ist es nun die Aufgabe von Experten solche Spektrogramme zu analysieren und auf potenzielle Fehler zu prüfen. Ähnlich wie bei Ärzten sind Ingenieure also dafür zuständig, Bilder zu Befunden und "Krankheiten"bzw. Fehler in Bauteilen zu finden.

3 Abstraktion für komplexe Datenstrukturen

Die dreidimensionalen Spektrogramme bieten die Grundlage für das angewandte ML Verfahren. Zuerst entsteht jedoch die Herausforderung, dass für jeden produzierten Motor ein Spektrogramm erzeugt wird und von Experten beurteilt werden muss. In der Serienfertigung hat dies zur Folge, dass zwar Prüfstände einen Großteil der anfallenden Daten bereits automatisiert auf Fehler untersuchten können, jedoch Experten immer noch mehrere tausend Spektrogramme manuell analysieren müssen. Deswegen werden Spektrogramme abstrahiert, um Fehlersymptome schneller erkennen zu können.

Da der Fokus dieses Kapitels nicht auf der Transformation der Daten, sondern der Unterstützung der Analyse durch Visualisierungen liegt, wird lediglich das Ergebnis dieses komplexen Transformations- und Abstraktionsprozesses beschrieben. In Zusammenarbeit mit Akustikexperten wurde hierzu die sogenannte "Hypermatrix" entwickelt. Zwei Hypermatrizen sind in Abbildung 3 dargestellt. Detaillierte

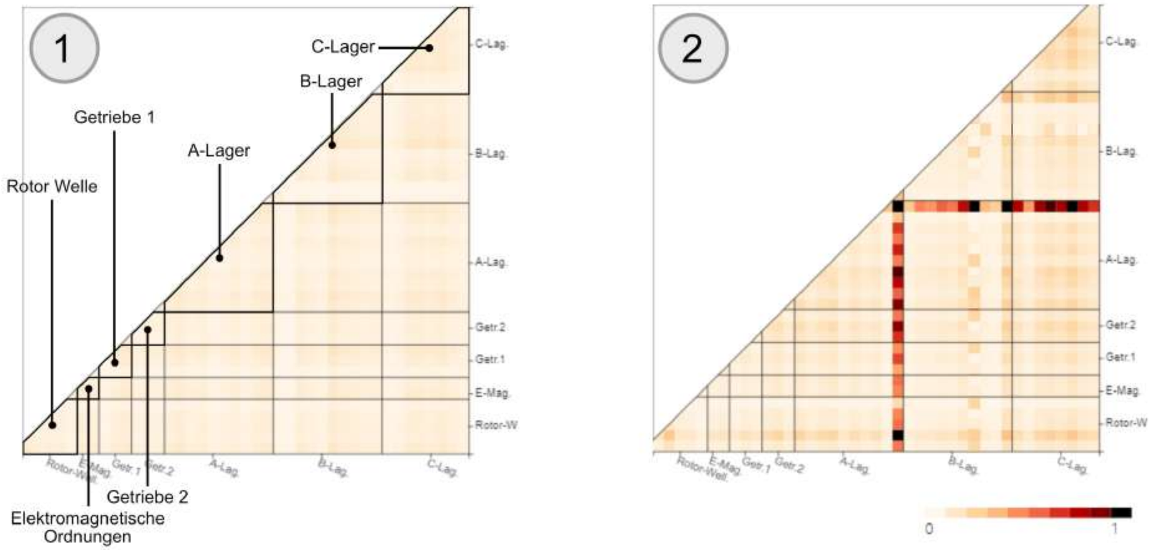


Abbildung 3: **Hypermatrizen von zwei Motoren:** Auf der X- und Y-Achse sind die Bauteile eines Motors abgebildet. Der Schweregrad einer Anomalie ist auf einer kontinuierlicheFarbskalale abgetragen, wobei Rot eine starke Anomalie und Hellorange keine Anomalie darstellen

Informationen zur Erstellung von Hypermatrizen sind in [2] und [3] zu finden. Wichtig für den Leser ist lediglich folgende Information.

1) Anhand der Motorordnungen aus dem im vorherigen Kapitel beschriebenen Spektrogramm kann der Standort eines zu analysierten Teils genau bestimmt werden (die Rotorwelle ist z. B. in der 1ten Motorordnung sichtbar). Ein Bauteil kann auch in mehreren Motorordnungen erkennbar sein. So ist die Rotorwelle in der 1ten, 2ten, 3ten und 4ten Motorordnung zu erkennen. Diese Bauteile sind auf der x- und y-Achse der Hypermatrix in Abbildung 3 abgetragen. Der Einfachheit halber, betrachten wir lediglich 41 Motorordnungen in dieser Hypermatrix, also 41 Spalten aus dem Spektrogramm. Diese 41 Ordnungen beschreiben die folgenden Bauteile: Rotorwelle, Elektromagnetische Ordnungen, 1tes Getriebe, 2tes Getriebe, A-Lager, B-Lager, C-Lager (Siehe 3 - 1).

2) Sobald ein Fehler auftritt, werden Kreuzstrukturen auf der Hypermatrix in Rottönen sichtbar. Das primäre Fehlersymptom ist an der Kreuzung der beiden Linien sichtbar (siehe 3 - 2). Sekundäre Fehlersymptome sind auf den Linien des Kreuzes sichtbar. Im Beispiel aus Abbildung 3 - 2 ist der Ursprung des dargestellten Fehlers im A-Lager. Dieser Fehler ist jedoch auch in den anderen Bauteilen eines Motors zu erkennen. Da die Matrix symmetrisch ist, wird nur eine Hälfte angezeigt.

Anhand der Hypermatrix fällt es augenscheinlich leichter Fehlerursachen zu erkennen als mithilfe eines Spektrogramms. Um den Experten die Analyse weiter zu erleichtern, gilt es nun im nächsten Schritt herauszufinden, ob erkannte Kreuzstrukturen in der Hypermatrix einmalige Ereignisse darstellen oder über mehrere Motoren ähnliche Strukturen erkennbar sind. Eine häufig auftretende Kreuzstruktur ist ein Indiz für wiederkehrende Fehlermuster, welche Experten nutzen können, um automatisierte Testverfahren in der Serienfertigung zu verbessern. Um Muster in Hypermatrizen zu erkennen, werden ähnliche Hypermatrizen gruppiert bzw. geclustert. Das Ziel hierbei ist es Experten ähnliche Hypermatrizen für ihre Befunden anzuzeigen, um somit deren Analyse zu beschleunigen. Das hier zugrunde liegende ML Verfahren wird im folgenden Kapitel erklärt.

4 Feature Engineering und Model Training

Um ähnliche Hypermatrizen zu gruppieren, extrahieren wir Merkmale aus der Hypermatrix jedes Motors. Der Vorgang der Feature Extraktion ist in Abbildung 4 dargestellt. Die Kombination der sieben Komponenten (z. B. Rotorwelle, elektromagnetische Ordnungen, Getriebe) des Motors führt zu 28 möglichen Kombinationen. In Abbildung 4 bezeichnen wir diese Kombination als Region (R). Aus jedem R wird die entsprechende Submatrix extrahiert. Aus jeder Submatrix extrahieren wir das

Maximum. Daraus ergibt sich ein 28x1-Feature Vektor. In unseren Anwendungsfall ist das Maximum ein sinnvolles Feature, da lautere Geräusche die Ursache für einen Fehler darstellen können. Je nach Use Case können auch unterschiedliche Features sinnvoll sein. Der 28x1-Feature Vektor stellt den Input für das ML Verfahren dar. Hierzu verwenden wir eine sogenannte “Self-Organizing Map (SOM)”[4]. SOMs sind unsupervised Verfahren auf Basis neuronaler Netze. Dabei wird ein Clustering mit einer Dimensionsreduktion kombiniert.

Grundsätzlich können neben SOMs auch andere Clustering-Techniken, z. B. KMeans (Die Bildung einer manuell bestimmten Anzahl von Gruppen ähnlicher Objekte) angewendet werden. In unserem Beispiel sind SOMs am besten geeignet, da Modellvorhersagen in einen überlappungsfreien rechteckigen Raum transformiert werden können. Resultierende Cluster können somit in einer Rasteransicht dargestellt werden. Dies ist besonders nützlich, da eine Rastervisualisierung die Ähnlichkeit zwischen Clustern bewahrt. Dies erleichtert in unserem Fall die Übersicht über Gruppen von Hypermatrizen und erleichtert deren Vergleichbarkeit.

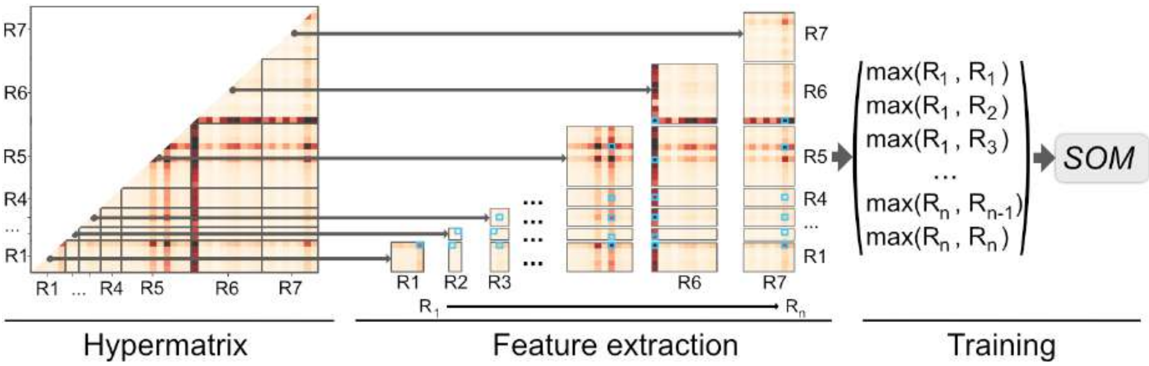


Abbildung 4: **Feature Engineering Prozess:** Aus jeder Hypermatrix, werden alle Submatrizen extrahiert. Anschließend wird aus jeder Submatrix, das Maximum extrahiert. Der daraus entstehende Vektor bildet den Input für das angewandte ML Verfahren.

5 Das Visual Analytics System: IRVINE

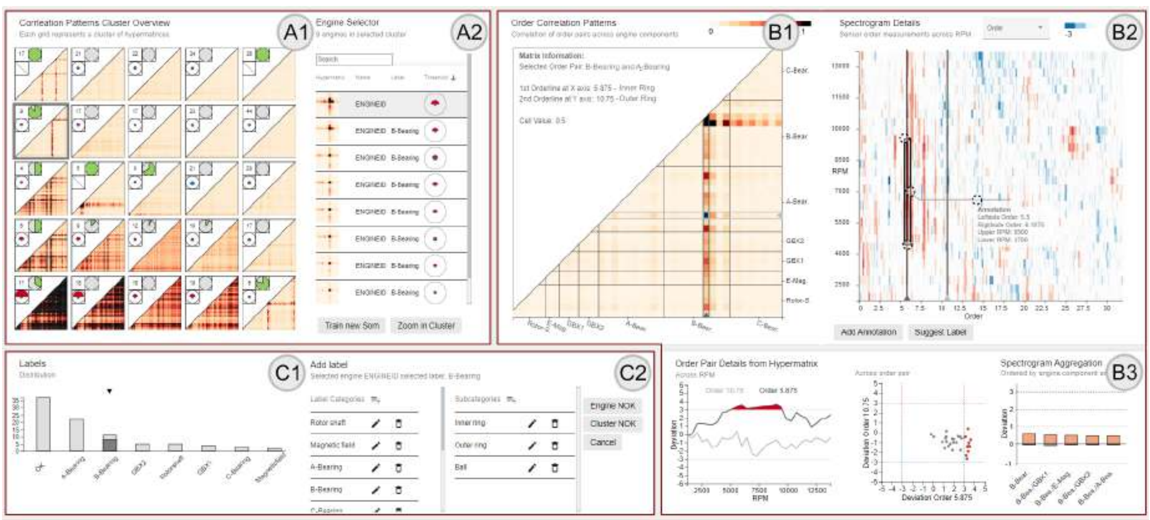
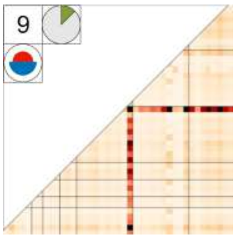


Abbildung 5: **Benutzerschnittstelle des VA Systems IRVINE:** In A sind die Cluster Ergebnisse der Hypermatrizen dargestellt. B ermöglicht eine detaillierte Analyse jedes einzelnen Motors, und C die Zuordnung von Labels zu Motoren und Clustern.

Im Folgenden beschreiben wir das vollständige IRVINE VA System. Als Input für das System dient die Hypermatrix und das Spektrogramm eines Motors. Abbildung 5 zeigt das User-Interface des Systems: In (A) sind die Resultate des SOM Clusterings dargestellt. Insgesamt sind 25 kleine Cluster als Quadrat in einer 5x5 Raster Visualisierung in A1 dargestellt. In jedem Cluster ist die gemittelte Hypermatrix abgebildet, welche aus allen Hypermatrizen im Cluster entsteht.



Auf jeder kleinen Hypermatrix sind außerdem drei kleine quadratische Visualisierungen zusätzlich dargestellt. Oben links befindet sich die Anzahl an Hypermatrizen in jeden Cluster. Rechts davon wird als Kreisdiagramm angezeigt, wie viele Hypermatrizen bereits gelabelt wurden. Unten links ist die aggregierte Abweichung der Residualwerte mit zwei Halbkreisen in rot und blau dargestellt. Je größer der Kreis, desto mehr weicht das Cluster vom Mittelwert aller Motoren ab. Dies ist ein Indiz dafür, dass hier besonders schlimme Fehler vorliegen. Je größer also die Kreise sind, desto interessanter ist ein Cluster für eine Detailanalyse durch einen Experten.

Sobald der Nutzer auf ein Cluster in A1 klickt, werden alle Hypermatrizen des ausgewählten Clusters in A2 in tabellarischer Form angezeigt. Links neben jeden ausgewählten Motor wird dessen Hypermatrix in einem niedrigen Pixelraum angezeigt. So können Nutzer auf einen Blick erkennen, ob die Motoren in einem Cluster ähnliche Hypermatrizen haben. Weiterhin wird in der Tabelle das Label des Motors angezeigt (falls vorhanden) und ebenfalls dessen aggregierte Abweichung der Residualwerte in Form von zwei Halbkreisen (wie schon beschrieben in rot und blau).

Nutzer, die eine Zeile aus der Tabelle in A2 auswählen, können die Hypermatrix in B1 und das Spektrogramm des Motors in B2 analysieren. Anhand der Hypermatrix können Experten nun einen Fehler schnell erkennen und anhand des Spektrogramms den Fehler validieren. Eine visuelle Verknüpfung der beiden Darstellungen erfolgt durch B3. Sobald Nutzer mit der Maus über B1 fahren, werden die entsprechenden Spalten aus B2 mit Linien angezeigt. Da jede Linie aus B2 auch als Kurve dargestellt werden kann, wird diese in B3 links dargestellt. Die Verteilung der beiden Linien wird als Streudiagramm dargestellt. Bauteile mit den höchsten Abweichungen werden als Balkendiagramm in B3 rechts dargestellt.

Nachdem ein Experte einen Fehler entdeckt hat, kann er in C2 den Motor einen Fehler zuordnen. Die Verteilung aller bisher vergeben Labels ist in C1 als Barchart dargestellt. Sollten alle Motoren in einem Cluster denselben Fehler haben, kann auch das ganze Cluster auf einen Schlag mit denselben Fehler gelabelt werden. Ganze Cluster in einem Zug zu labeln ist vor allem hilfreich, wenn besonders viele Motoren mit ähnlichen Mustern vorliegen und nicht jeder einzelne Motor im Detail analysiert werden kann.

Nachdem ein Label in C2 vergeben wurde, können Experten auch die Ursache des Fehlers in B2 annotieren. Beispielhafte Annotationen sind in Abbildung 6 dargestellt.

Das System IRVINE erlaubt es somit Experten nicht nur einfacher und schneller Motoren zu analysieren, sondern schafft auch eine Plattform anhand der automatisch Expertenwissen in Form von Labels und Annotationen gespeichert wird. Dieses Wissen kann nun für verschiedenste Zwecke verwendet werden. Beispielsweise kann IRVINE basierend auf bereits gelabelten und annotierten Motoren Vorschläge liefern, welche Regionen im Spektrogramm besonders relevant für einen potenziellen Fehler sind.

Das mit IRVINE externalisierte Expertenwissen hilft auch dabei, neue Mitarbeiter in der Analyse von Spektrogrammen zu schulen, da diese sich gezielt Fehlerbilder und deren Ursachen anhand der Annotationen anschauen kann. Weiterhin können Data Scientisten die gelabelte Daten nutzen, um noch bessere ML Modelle zu trainieren, welche den Produktionsprozess optimieren können. Anhand der Labels können Modelle erkennen, wo ein Fehler aufgetreten ist und anhand der Annotationen die Ursache besser verstehen. IRVINE schafft damit einen wichtigen Baustein für eine weiterführende, KI-basierte Automatisierung des Prozesses.

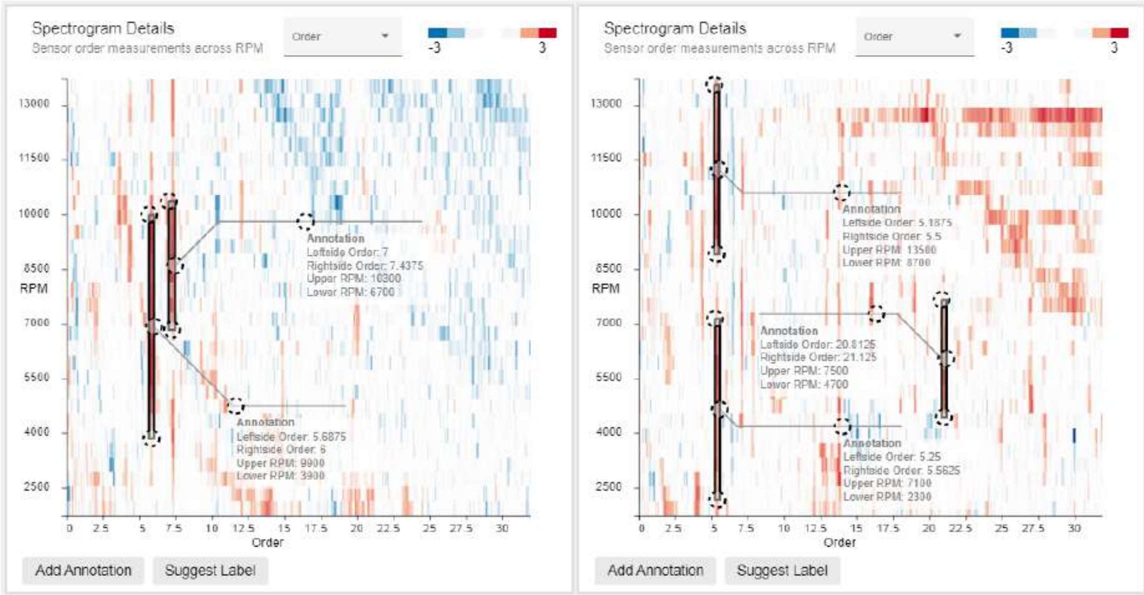


Abbildung 6: **Visuelle Annotationen von Motoren mit unterschiedlichen Fehlerbildern:** Annotationen können per "Drag und Drop" als Polygon einem Spektrogramm zugeordnet werden. Für jedes Spektrogramm werden die Annotationen gespeichert.

Danksagungen

Wir danken dem BMW Group KI Team und Fachstellen, insbesondere Bernhard Pflugfelder, Kathrin Berkel und Benny Kneissl. Ein weiterer Dank geht an die Entwickler des VA Systems IRVINE, insbesondere Jakob Bonart, Dominik Jäckle, Ute Schmid, Tobias Schreck und Jürgen Bernard.

Literatur

[1] James J. Thomas and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[2] Joscha Eirich, Jakob Bonart, Dominik Jäckle, Michael Sedlmair, Ute Schmid, Kai Fischbach, Tobias Schreck, and Jürgen Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. *IEEE Transactions on Visualization and Computer Graphics*, 07 2021.

[3] Jakob Bonart, Joscha Eirich, Philipp Hümmer, and Welf-Guntram Drossel. Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms. 08 2021.

[4] T. Kohonen, M. R. Schroeder, and T. S. Huang. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 3rd edition, 2001.

Bibliography

[1] W. J. Abernathy. *The productivity dilemma: roadblock to innovation in the automobile industry*. Johns Hopkins University Press, Baltimore, 1978.

[2] M. Alavi and D. E. Leidner. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1):107–136, 2001.

[3] S. Allard. Knowledge creation. In *Handbook on knowledge management. 1: Knowledge matters*, pp. 367–380. Springer, New York, 2004.

[4] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[5] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. Viewing Visual Analytics as Model Building. *COMPUTER GRAPHICS FORUM*, 37(6):275–299, Sept. 2018. doi: 10.1111/cgf.13324

[6] L. Argote. *Organizational Learning: Creating, Retaining, and Transferring Knowledge*. Springer, USA, 2dn ed., 2013.

[7] L. Argote and E. Miron-Spektor. Organizational learning: From experience to knowledge. *Organization Science*, 22(5):1123–1137, 2011.

[8] C. Argyris and D. A. Schön. *Organizational learning*. Addison-Wesley OD series. Addison-Wesley Pub. Co, Reading, Mass, 1978.

[9] A. Barsky, T. Munzner, and R. Kincaid. Cerebral: Visualizing multiple experimental conditions on a graph with biological context. *IEEE transactions on visualization and computer graphics*, 14:1253–60, 11 2008.

[10] M. J. Bates. Information and knowledge: an evolutionary framework for information science (invited paper). *Inf. Res.*, 10, 2005.

[11] S. Berg, D. Kutra, T. Kroeger, C. Straehle, B. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Köthe, F. Hamprecht, and A. Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16:1–7, 2019.

[12] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An experimental study. *Transactions on Visualization and Computer Graphics*, PP:1–1, 2017.

[13] J. Bonart, J. Eirich, P. Hümmer, and W.-G. Drossel. Identification of anomalies in highly-integrated electric drives by secondary excitation mechanisms. In *IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, 08 2021.

[14] M. Booshehrian, T. Möller, R. Peterman, and T. Munzner. Vismon: Facilitating analysis of trade-offs, uncertainty, and sensitivity in fisheries management decision making. *Computer Graphics Forum*, 31:1235–1244, 06 2012.

[15] S. Card and J. Mackinlay. The structure of the information visualization design space. In *Proceedings of VIZ '97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium*, pp. 92–99, 1997.

[16] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision To Think*. 01 1999.

[17] V. W. Chao, Y. Wu, L. Zhao, V. Tsai, and C.-H. Chen. Seismologically determined bedload flux during the typhoon season. *Scientific Reports*, 5, 2015.

[18] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver. Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1):12–19, 2009. doi: 10.1109/MCG.2009.6

[19] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. Paved: Pareto front visualization for engineering design. *Computer Graphics Forum*, 39:405–416, 2020.

[20] S. Clegg and T. Clarke. Intelligence: Organizational. In *International Encyclopedia of the Social & Behavioral Sciences*, pp. 7665–7670. Elsevier, 2001.

[21] K. Dalkir. *Knowledge management in theory and practice*. MIT Press, Cambridge, Mass, 2nd ed ed., 2011.

[22] P. Darke, G. Shanks, and M. Broadbent. Successfully completing case study research: combining rigour, relevance and pragmatism. *Information Systems Journal*, 8(4):273–289, Oct. 1998.

[23] T. H. Davenport. Business Intelligence and Organizational Decisions:. *International Journal of Business Intelligence Research*, 1(1):1–12, Jan. 2010.

[24] T. H. Davenport and L. Prusak. *Working knowledge: how organizations manage what they know*. Harvard Business School Press, Boston, Mass, 1998.

[25] I. R. Edvardsson. Knowledge management in smes: the case of icelandic firms. *Knowledge Management Research & Practice*, 4(4):275–282, 2006.

[26] I. R. Edvardsson. Is knowledge management losing ground? developments among icelandic smes. *Knowledge Management Research & Practice*, 7(1):91–99, 2009.

[27] J. Eirich. Visual analytics for iot data from large-scale manufacturing processes. *American Conference on Information Systems*, 08 2022.

[28] J. Eirich, J. Bonart, D. Jäckle, M. Sedlmair, U. Schmid, K. Fischbach, T. Schreck, and J. Bernard. Irvine: A design study on analyzing correlation patterns of electrical engines. *IEEE Transactions on Visualization and Computer Graphics*, pp. 11–21, 07 2021.

[29] J. Eirich and F.-P. Diana. The life cycle of data labels in organizational learning: A case study of the automotive industry. In *European Conference on Information Systems*, 04 2022.

[30] J. Eirich, D. Jäckle, T. Schreck, J. Bonart, O. Posegga, and K. Fischbach. Vima: Modeling and visualization of high dimensional machine sensor data leveraging multiple sources of domain knowledge. In *Visualization in Data Science at IEEE VIS*, 09 2020.

[31] J. Eirich, D. Jäckle, M. Sedlmair, C. Wehner, U. Schmid, J. Bernard, T. Schreck, and J. Bernard. Manknowvis: How to support different user groups in contextualizing and leveraging knowledge repositories. *IEEE Transactions on Visualization and Computer Graphics*, 07 2022.

[32] J. Eirich, D. Jäckle, S. Werrlich, and T. Schreck. Visual analytics in organizational knowledge creation: A case study. In *European Conference on Information Systems*, 04 2021.

[33] J. Eirich, G. Koutroulis, B. Mutlu, D. Jäckle, R. Kern, T. Schreck, and J. Bernard. Manex: The visual analysis of measurements for the assessment of errors in electrical engines. *IEEE Computer Graphics and Applications*, pp. 1–13, 03 2022.

[34] J. Eirich, A. Luckow, and D. K. Visual analytics zur optimierung von pruefprozessen in der serienfertigung elektrischer motoren. In *Wie funktionier Data Science?*, pp. 1–7. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

[35] J. Eirich, M. Münch, D. Jäckle, M. Sedlmair, J. Bonart, and T. Schreck. Rfx: A design study for the interactive exploration of a random forest to enhance testing procedures for electrical engines. *Computer Graphics Forum*, pp. 1–14, 03 2022.

[36] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Mag.*, 17:37–54, 1996.

[37] P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 92–103, 2017. doi: 10.1109/VAST.2017.8585498

[38] J. L. Flanagan. *Speech analysis; synthesis and perception*. Springer, 2nd ed., 1972.

[39] T. Fujiwara, S. Chandrasegaran, M. Brundage, T. Sexton, A. Dima, and K. Ma. A visual analytics approach for the diagnosis of heterogeneous and multidimensional machine maintenance data. In *Proceedings - 2021 IEEE 14th Pacific Visualization Symposium, PacificVis 2021*, IEEE Pacific Visualization Symposium, pp. 196–205. IEEE, United States, 2021.

[40] M. Gashi, B. Mutlu, S. Lindstaedt, and S. Thalmann. Decision support for multi-component systems: visualizing interdependencies for predictive maintenance. In *Proceedings of the 55th Hawaii International Conference on System Sciences*, Jan. 2022.

[41] R. M. Grant. Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science*, 7(4):375–387, Aug. 1996.

[42] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From visual exploration to storytelling and back again. *Computer Graphics Forum (EuroVis '16)*, 35(3):491–500, jun 2016.

[43] T. M. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8(1):1–13, 2009.

[44] T. Greensmith, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8:1–13, 03 2009. doi: 10.1057/ivs.2008.28

[45] E. Grimsdottir and I. R. Edvardsson. Knowledge management, knowledge creation, and open innovation in icelandic smes. *SAGE Open*, 8(4):2158244018807320, 2018.

[46] T. Grønsund and M. Aanestad. Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2):101614, June 2020.

[47] H. Guo, S. R. Gomez, C. Ziemkiewicz, and D. H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):51–60, jan 2016.

[48] M. T. Hansen, N. Nohria, and T. Tierney. What's your strategy for managing knowledge? *Harvard Business Review*, 77(2), 1999.

[49] F. Heimerl, S. Koch, H. Bosch, and T. Ertl. Visual classifier training for text document retrieval. *Transactions on Visualization and Computer Graphics*, 18(12):2839–2848, 2012.

[50] A. Hevner, A. R, S. March, S. T, Park, J. Park, Ram, and Sudha. Design science in information systems research. *Management Information Systems Quarterly*, 28:75–, 03 2004.

[51] S. Hughes. Medical ultrasound imaging. *Physics Education*, 36:468, 2001.

[52] K. Iyengar, J. R. Sweeney, and R. Montealegre. Information Technology Use as a Learning Mechanism: The Impact of IT Use on Knowledge Transfer Effectiveness, Absorptive Capacity, and Franchisee Performance. *MIS Quarterly*, 39(3):615–641, Mar. 2015.

[53] A. Jashapara. *Knowledge management: an integrated approach*. Financial Times Prentice Hall, Harlow Munich, 2nd ed ed., 2011.

[54] N. Jekic, B. Mutlu, M. Schreyer, S. Neubert, and T. Schreck. Similarity measures for visual comparison and retrieval of test data in aluminum production. In C. Hurter, H. Purchase, J. Braz, and K. Bouatouch, eds., *IVAPP*, VISIGRAPP 2021 - Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 210–218. SciTePress 2013, 2021.

[55] N. Jekic, M. Schreyer, S. Neubert, B. Mutlu, and T. Schreck. Visual data analysis of production quality data for aluminum casting. In T. Bui, ed., *Proceedings of the 54th Annual Hawaii International Conference on System Sciences, HICSS 2021*, Proceedings of the Annual Hawaii International Conference on System Sciences, pp. 1487–1495. ScholarSpace, 2021. 54th Annual Hawaii International Conference on System Sciences : HICSS 2021 ; Conference date: 04-01-2021 Through 08-01-2021.

[56] G. Kane and M. Alavi. Information technology and organizational learning: An investigation of exploration and exploitation processes. In *Organization Science*, vol. 18, 01 2005.

[57] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual Analytics: Scope and Challenges. In S. J. Simoff, M. H. Böhlen, and A. Mazeika, eds., *Visual Data Mining*, vol. 4404, pp. 76–90. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-71080-6_6

[58] S. Kim, T. Mukhopadhyay, and R. Kraut. When does repository kms use lift performance? the role of alternative knowledge sources and task environments. *MIS Quarterly: Management Information Systems*, 40(1):133–156, 3 2016.

[59] K. Koffka. *Principles of Gestalt psychology*. Principles of Gestalt psychology. Harcourt, Brace, Oxford, England, 1935. Pages: 720.

[60] T. Kohonen. Essentials of the self-organizing map. *Neural Networks*, 37:52–65, 2013.

[61] T. Kohonen, M. R. Schroeder, and T. S. Huang. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 3rd ed., 2001.

[62] B. L. Kovitz. *Practical software requirements: a manual of content and style*. Manning, Greenwich, Conn, 1999.

[63] A. Kusiak. Smart manufacturing. *International Journal of Production Research*, 56(1-2):508–517, 2018.

[64] H. Kyriakou, J. V. Nickerson, and G. Sabnis. Knowledge Reuse for Customization: Metamodels in an Open Design Community for 3D Printing. *MIS Quarterly*, 41(1):315–332, 1 2017.

[65] T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, and A. Rind. Towards a Concept how the Structure of Time can Support the Visual Analytics Process. In S. Miksch and G. Santucci, eds., *EuroVA 2011: International Workshop on Visual Analytics*. The Eurographics Association, 2011.

[66] R. Larsson, L. Bengtsson, K. Henriksson, and J. Sparks. The Interorganizational Learning Dilemma: Collective Knowledge Development in Strategic Alliances. *Organization Science*, 9(3):285–305, 1998.

[67] S. Lebovitz, N. Levina, and H. Lifshitz-Assa. Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, 45:1501–1526, Sept. 2021.

[68] J. Lewis and J. Sauro. Item benchmarks for the system usability scale. 13:158–167, 05 2018.

[69] U. Lichtenthaler and E. Lichtenthaler. A Capability-Based Framework for Open Innovation: Complementing Absorptive Capacity. *Journal of Management Studies*, 46(8):1315–1338, 12 2009.

[70] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2498–2507, dec 2011.

[71] A.-P. Lohfink, S. D. Duque Anton, H. Leitte, and C. Garth. Knowledge rocks: Adding knowledge assistance to visualization systems. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1117–1127, 2022.

[72] B. Lorensen. On the death of visualization. In *Position Papers NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges*, vol. 1, p. 5, 2004.

[73] A. Luckow, J. Eirich, and D. K. Kuenstliche intelligenz: Anwendungen und werkzeuge in die automobilindustrie. In *Wie funktionier Data Science?*, pp. 1–11. Springer, Heidelberg Platz 3, 14197 Berlin, 2022.

[74] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, apr 1986.

[75] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 20:1633–1642, 12 2014.

[76] A. Maier, T. Tack, and O. Niggemann. Visual anomaly detection in production plants. In *International Conference on Informatics in Control, Automation and Robotics*, vol. 1, 07 2012.

[77] J. G. March. Exploration and Exploitation in Organizational Learning. *Organization Science*, 2(1):71–87, 1991.

[78] R. Maruta. The creation and management of organizational knowledge. *Know.-Based Syst.*, 67:26–34, sep 2014.

[79] A. Merono-Cerdan, C. López-Nicolás, and R. Sabater-Sánchez. Knowledge management strategy diagnosis from km instruments use. *J. Knowledge Management*, 11:60–72, 04 2007.

[80] S. Miksch and W. Aigner. A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data. *Computers and Graphics, Special Section on Visual Analytics*, 38:286–290, 2014.

[81] S. Miksch, H. Leitte, and M. Chen. *Knowledge-Assisted Visualization and Guidance*, pp. 61–85. Springer International Publishing, 08 2020.

[82] J. Moehrmann, S. Bernstein, T. Schlegel, G. Werner, and G. Heidemann. Improving the usability of hierarchical representations for interactively labeling large image data sets. In *Human computer interactions*, p. 618–627. Springer, Berlin, Heidelberg, 2011.

[83] D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):438–448, 2019.

[84] T. Munzner. A nested model for visualization design and validation. *Visualization and Computer Graphics*, 15:921 – 928, 2010.

[85] K. Nie, P. Baltzer, B. Preim, and G. Mistelbauer. Knowledge-Assisted Comparative Assessment of Breast Cancer using Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Computer Graphics Forum*, 2020.

[86] S. Nobarany, M. Haraty, and B. Fisher. Facilitating the reuse process in distributed collaboration: A distributed cognition approach. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, p. 1223–1232. Association for Computing Machinery, New York, NY, USA, 2012.

[87] I. Nonaka, M. Kodama, A. Hirose, and F. Kohlbacher. Dynamic fractal organizations for promoting knowledge-based transformation – a new paradigm for organizational theory. *European Management Journal*, 32(1):137–146, 2014.

[88] I. Nonaka and H. Takeuchi. *The knowledge-creating company: How japanese companies create the dynamics of innovation*. Oxford University Press, New York, 1995.

[89] I. Nonaka, R. Toyama, and N. Konno. SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation. *Long Range Planning*, 33(1):5–34, Feb. 2000.

[90] B. T. Pentland. Information systems and organizational learning: The social epistemology of organizational knowledge systems. *Accounting, Management and Information Technologies*, 5(1):1–21, Jan. 1995.

[91] S. Pinker and R. Feedle. *A theory of graph comprehension*, pp. 73–126. 01 1990.

[92] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, 01 2005.

[93] J. Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, 1983. doi: 10.1109/TSMC.1983.6313160

[94] D. Ren, M. Brehmer, B. Lee, T. Höllerer, and E. K. Choe. Chartaccent: Annotation for data-driven storytelling. *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 230–239, 2017.

[95] J. Rowley. The wisdom hierarchy: representations of the dikw hierarchy. *Journal of Information Science*, 33(2):163–180, 2007.

[96] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1:206–215, 2019.

[97] T. Ruppert, M. Staab, A. Bannach, H. Lücke-Tieke, J. Bernard, A. Kuijper, and J. Kohlhammer. Visual interactive creation and validation of text clustering workflows to explore document collections. *Electronic Imaging*, 2017(1):46–57, 2017.

[98] D. Sacha, M. Kraus, J. Bernard, M. Behrisch, T. Schreck, Y. Asano, and D. A. Keim. Somflow: Guided exploratory cluster analysis with self-organizing maps and analytic provenance. *Transactions on Visualization and Computer Graphics*, 24(1):120–130, 2018.

[99] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, Jan. 2016. doi: 10.1109/TVCG.2015.2467591

[100] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014.

[101] T. Sanislav and L. Miclea. Cyber-physical systems - concept, challenges and research areas. *Control Engineering and Applied Informatics*, 14:28–33, 01 2012.

[102] M. Sedlmair, C. Bernhold, D. Herrscher, S. Boring, and A. Butz. Mostvis: An interactive visualization supporting automotive engineers in most catalog exploration. In *2009 13th International Conference Information Visualisation*, pp. 173–182, 2009. doi: 10.1109/IV.2009.95

[103] M. Sedlmair, A. Frank, T. Munzner, and A. Butz. Relex: Visualization for actively changing overlay network specifications. *Visualization and Computer Graphics*, 18:2729–2738, 2012.

[104] M. Sedlmair, P. Isenberg, D. Baur, M. Mauerer, C. Pigorsch, and A. Butz. Cardiogram: Visual analytics for automotive engineers. In *Human Factors in Computing Systems*, pp. 1727–1736, 2011.

[105] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics*, 18:2431–2440, 2012.

[106] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages*, p. 336. IEEE Computer Society, USA, 1996.

[107] A. Shollo and R. D. Galliers. Towards an Understanding of the Role of Business Intelligence Systems in Organisational Knowing. *Information Systems Journal*, 26(4):339–367, 2016.

[108] A. Smith and C. Zeithaml. Garbage Cans and Advancing Hypercompetition: The Creation and Exploitation of New Capabilities and Strategic Flexibility in Two Regional Bell Operating Companies. *Organization Science*, 7:388–399, 1996.

[109] T. Sobral, T. Galvão, and J. Borges. An ontology-based approach to knowledge-assisted integration and visualization of urban mobility data. *Expert Systems with Applications*, 150:113260, 2020.

[110] M. Someren, Y. Barnard, and J. Sandberg. *The Think Aloud Method - A Practical Guide to Modelling CognitiveProcesses*. Academic Press, 1994.

[111] H. Stitz, S. Gratzl, H. Piringer, T. Zichner, and M. Streit. Knowledgepearls: Provenance-based visualization retrieval. *IEEE Transactions on Visualization and Computer Graphics (VAST '18)*, 25(1):120–130, 2018. doi: 10.1109/TVCG.2018.2865024

[112] T. Sturm, J. Gerlach, L. Pumplun, N. Mesbah, F. Peters, C. Tauchert, N. Nan, and P. Buxmann. Coordinating Human and Machine Learning for Effective Organizational Learning. *MIS Quarterly*, 45(3):1581–1602, Mar. 2021.

[113] D. Sun, R. Huang, Y. Chen, Y. Wang, J. Zeng, M. Yuan, T.-C. Pong, and H. Qu. Planningvis: A visual analytics approach to production planning in smart factories. *IEEE transactions on visualization and computer graphics*, PP, 08 2019.

[114] J. Suschnigg, B. Mutlu, A. Fuchs, V. Sabol, S. Thalmann, and T. Schreck. Exploration of anomalies in cyclic multivariate industrial time series data for condition monitoring. In *Proc. 3rd Int. Workshop on Big Data Visual Exploration and Analytics EDBT/ICDT 2020*, 2020.

[115] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[116] K. Trantopoulos, G. von Krogh, M. W. Wallin, and M. Woerter. External Knowledge and Information Technology: Implications for Process Innovation Performance. *MIS Quarterly*, 41(1):287–300, Mar. 2017.

[117] V. Vaishnavi and W. Kuechler. *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, p. 226. 10 2007.

[118] E. van den Broek, A. Sergeeva, and M. Huysman Vrije. When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, 45(3):1557–1580, Sept. 2021.

[119] J. van Wijk. The Value of Visualization. In *VIS 05. IEEE Visualization, 2005.*, pp. 79–86. IEEE, Minneapolis, MN, USA, 2005.

[120] J. van Wijk. Views on visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):421–432, 2006.

[121] D. Vera and M. Crossan. Organizational learning and knowledge management: toward an integrative framework. *Handbook of Organizational Learning and Knowledge Management*, pp. 122–141, Jan. 2003.

[122] M. Vukovic, V. Dhanoa, M. Jäger, C. Walchshofer, J. Küng, P. Krahwinkler, B. Mutlu, and S. Thalmann. A forecasting model-based discovery of causal links of key influencing performance quality indicators for sinter production improvement. In *2020 AISTech Conference Proceedings*, pp. 2028–2038, 2020.

[123] M. Wagner, D. Slijepcevic, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner. KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 25(3):1528–1542, Mar. 2019. doi: 10.1109/TVCG.2017.2785271

[124] M. Waldner, T. Geymayer, D. Schmalstieg, and M. Sedlmair. Linking unstructured evidence to structured observations. *Information Visualization*, 20(1):47–65, 2021.

[125] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019.

[126] C. Wang and K.-L. Ma. Information and knowledge assisted analysis and visualization of large-scale data. In *2008 Workshop on Ultrascale Visualization*, pp. 1–8, 2008.

[127] X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang. Defining and applying knowledge conversion processes to a visual analytics system. *Comput. Graph.*, 33:616–623, 2009.

[128] K. Weick, K. Sutcliffe, and D. Obstfeld. Organizing and the process of sensemaking. *ORGANIZATION SCIENCE*, 16:409–421, 07 2005. doi: 10.1287/orsc.1050.0133

[129] S. Werrlich, P.-A. Nguyen, A.-D. Daniel, C. E. F. Yanez, C. Lorber, and G. Notni. Design recommendations for hmd-based assembly training tasks. In *SmartObjects@ CHI*, pp. 58–68, 2018.

[130] M. Wörner, M. Metzger, and T.Ertl. Dataflow-based Visual Analysis for Fault Diagnosis and Predictive Maintenance in Manufacturing. In M. Pohl and H. Schumann, eds., *EuroVis Workshop on Visual Analytics*, 2013.

[131] D. Wu, X. Wang, J. Su, B. Tang, and S. Wu. A labeling method for financial time series prediction based on trends. *Entropy*, 22(10), 2020.

[132] L. Xiao, J. Gerth, and P. Hanrahan. Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation. In *2006 IEEE Symposium On Visual Analytics And Technology*, pp. 107–114. IEEE, Baltimore, MD, USA, Oct. 2006.

[133] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *Transactions on Visualization and Computer Graphics*, 23:1–1, 01 2016.

[134] H. Yang, S. Kumara, S. T. Bukkapatnam, and F. Tsung. The internet of things for smart manufacturing: A review. *IISE Transactions*, 51(11):1190–1216, 2019.

[135] R. K. Yin. *Case study research: design and methods*. Number v. 5 in Applied social research methods series. Sage Publications, Thousand Oaks, Calif, 3rd ed ed., 2003.

[136] R. K. Yin. *Case study research: design and methods*. Number v. 5 in Applied social research methods series. Sage Publications, Thousand Oaks, Calif, 5th ed ed., 2014.

[137] J. Young and a. O. M. C. Safari. *Personal Knowledge Capital*. Chandos Publishing, 2012.

[138] J. Zhao, M. Glueck, P. Isenberg, F. Chevalier, and A. Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Transactions on Visualization and Computer Graphics*, PP, 08 2017.

[139] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22:419–435, 2019.