# INTEGRATING KNN WITH FOUNDATION MODELS FOR ADAPTABLE AND PRIVACY-AWARE IMAGE CLASSIFICATION

*Sebastian Doerrich** *Tobias Archut** *Francesco Di Salvo* *Christian Ledig*

xAILab, University of Bamberg, Germany

## ABSTRACT

Traditional deep learning models implicitly encode knowledge limiting their transparency and ability to adapt to data changes. Yet, this adaptability is vital for addressing user data privacy concerns. We address this limitation by storing embeddings of the underlying training data independently of the model weights, enabling dynamic data modifications without retraining. Specifically, our approach integrates the $k$-Nearest Neighbor ($k$-NN) classifier with a vision-based foundation model, pre-trained self-supervised on natural images, enhancing interpretability and adaptability. We share open-source implementations of a previously unpublished baseline method as well as our performance-improving contributions. Quantitative experiments confirm improved classification across established benchmark datasets and the method's applicability to distinct medical image classification tasks. Additionally, we assess the method's robustness in continual learning and data removal scenarios. The approach exhibits great promise for bridging the gap between foundation models' performance and challenges tied to data privacy. The source code is available at github.com/TobArc.

***Index Terms*—** $k$-NN classifier, continual learning, transfer learning, few-shot classification, explainability

## 1. INTRODUCTION

Deep learning has exhibited significant success in diverse domains, notably in natural language processing [1, 2] and image classification [3, 4, 5], driven by the evolution of increasingly sophisticated models. These models, empowered by substantial computational resources, excel in capturing intricate patterns and implicit representations within their parameters [6]. However, the inherent limitation of tying knowledge exclusively to model weights introduces a significant drawback. The opacity of this knowledge restricts efficient information retrieval [7] and raises concerns about data usage rights and privacy [8]. This challenge is accentuated by evolving regulations, such as the European Union's *right to erasure ('right to be forgotten')* (Article 17 of the General Data Protection Regulation (GDPR) [9]), empowering users to revoke data usage rights promptly.

Updating knowledge in deep learning models, involving tasks such as addition, deletion, or modification of information, currently necessitates comprehensive retraining or fine-tuning [10]. This process incurs substantial computational expenses and proves cumbersome, particularly in sensitive sectors like healthcare. The predominant paradigm of exclusive data storage within model parameters lacks adaptability, especially when users exercise their right to update or delete personal data. This leads to exponential costs and the risk of catastrophic forgetting in continual learning scenarios [11], rendering these models unpracticable at best, infeasible, or irresponsible at worst.

In response to these challenges, our research is inspired by Nakata et al.'s solution [7], which deviates from the conventional approach of storing knowledge solely in model parameters. We advocate for storing comprehensive training dataset knowledge, including image feature representations and labels, in an external dynamic repository. This approach enables seamless addition, deletion, or modification of data without necessitating model retraining. Integrating the classical $k$-Nearest Neighbor ($k$-NN) classifier [12] with the robust and discriminative feature spaces of foundation models, pre-trained in a self-supervised manner on natural images, enhances interpretability and adaptability. Our contributions encompass:

- Open-source implementation including an independent performance validation of Nakata et al.s work for which there is currently no public implementation available.
- Advancing the method's performance by incorporating recent foundation models and a more flexible data storage system, enabling few-shot adaptation for medical image analysis.
- Quantitative confirmation that the method addresses data privacy concerns by facilitating task-incremental learning as well as allowing for data removal in sensitive healthcare scenarios without compromising model performance.

## 2. RELATED WORK

Foundation models, exemplified by Transformer [13] and Vision Transformer [3], demand extensive training on large-scale datasets to excel in tasks like natural language processing [2] or image generation [14]. Self-supervised contrastive

* These authors contributed equally to this work.

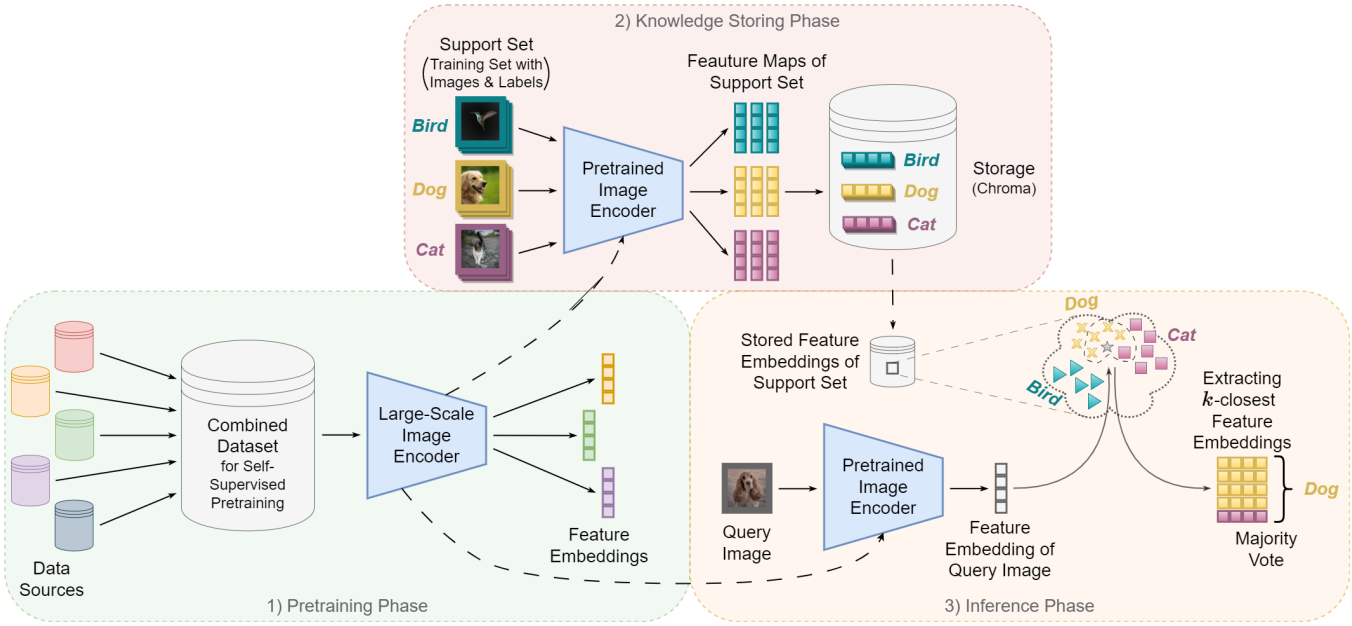



**Fig. 1**: During pretraining (1), the image encoder is trained to extract representative features. The knowledge-storing phase (2) utilizes the pre-trained (now frozen) encoder to extract and store task-relevant knowledge from the training data. During inference (3), that knowledge allows the classification of query images through majority voting on the top-$k$ similar embeddings.

methods, such as CLIP [4] (on image-text pairs) and DINOv2 [15] (exclusively on image pairs), enable training without intensive labeling, yet fine-tuning on annotated datasets is often necessary to optimize results. However, fine-tuning poses a risk of catastrophic forgetting in continual learning scenarios. Model-based approaches like GEM [16], ER [17], and iCaRL [18] mitigate this but lock all knowledge in the model's weights, limiting information retrieval and modification possibilities. In contrast, $k$-NN methods, previously employed for representation learning evaluation [5] or noise reduction [19], proved promising for enhancing knowledge retrieval as shown by RETRO [20] in auto-regressive language models.

Nakata et al. [7] combine these foundations, employing a $k$-NN classifier in a three-phase methodology, which involves pretraining on natural images, knowledge storage through feature map extraction, and inference based on $k$-NN retrieval. This eliminates the need for fine-tuning and exhibits efficacy in continual learning scenarios. Our work extends this innovation by integrating the $k$-NN classifier with recent vision-based foundation models. Specifically, we propose to extract image features with DINOv2 [15], preserving robustness and adaptability across diverse scenarios while enhancing classification. Additionally, separating computation from storage ensures flexible knowledge management, addressing data privacy concerns in particular. We further extend the method and quantitatively confirm its applicability to the sensitive healthcare domain by demonstrating uncompromised performance in challenging task-incremental learning and seamless data removal scenarios.

## 3. METHOD

### 3.1. Overview

Our approach utilizes a three-phase structure as depicted in Figure 1 and further outlined below:

**Pretraining Phase** Initially, a foundation model is pretrained on a large-scale dataset, accommodating unlabeled or noisily labeled images to obviate upfront labeling costs. The focus is on extracting generic, vision task-independent features crucial for robust and reliable $k$-NN performance. The choice of an image encoder trained on a diverse dataset becomes imperative for effective segregation of feature embeddings, facilitating robust generalization across datasets.

**Knowledge-Storing Phase** In the knowledge-storing phase, the pre-trained image encoder captures feature embeddings from the training set (support set) which are subsequently stored along with the corresponding labels in an external database. This way, task-relevant knowledge is kept separate from the encoder's weights, adhering to continual learning paradigms and privacy regulations. This design allows seamless addition, modification, and deletion of samples.

**Inference Phase** During inference, the pre-trained image encoder generates a feature embedding for a given query image. Top-$k$ similar feature embeddings are retrieved from the external database using cosine similarity as the distance metric. We use cosine similarity due to its robustness in capturing scale-invariant angular relationships between vectors, making it particularly effective for measuring similarity in multi-dimensional data representations [21]. Classification of

the query image is determined through a majority vote on the labels associated with the top-$k$ similar feature embeddings, enabling efficient classification without encoder retraining.

### 3.2. Backbone architecture

Differing from Nakata et al. [7], we opt for the DINOv2 [15] backbone over CLIP [4] to enhance the robustness of our method. DINOv2 employs self-supervised contrastive training on 142 million distinct images from curated and uncurated data sources, emphasizing high-quality feature representation by minimizing the distance of similar objects and maximizing the distance of distinct ones. We choose DINOv2 Large[1][3] with $14 \times 14$ patches and 1024-sized image embeddings over its Base[2][3] version, to increases model capacity and feature representation (304M parameters vs. 87M parameters).

### 3.3. Knowledge storage

Nakata et al. [7] require loading both the image encoder model and all stored feature embeddings into a singular processing unit's memory, resulting in significant computational demands, especially for large support sets. This imposes a continuous need for substantial computational resources. To mitigate this challenge, our strategy involves the active separation of storage from computational processes. Utilizing Chroma [22], an open-source, in-memory embedding database, we ensure efficient storage and retrieval of feature embeddings. It's essential to note that alternative vector database solutions could be employed including highly efficient approximate nearest neighbor search algorithms.

## 4. EXPERIMENTS AND RESULTS

We first evaluate the choice of backbone in terms of our method's classification ability on natural images. We further assess the adaptability of our approach to image classification tasks in the medical domain including its ability for task-incremental learning and its potential for seamless removal of sensitive, task-relevant data without seriously compromising performance. To this end, we utilize a comprehensive set of distinct datasets comprising natural images, such as CIFAR-10 [23], CIFAR-100 [23], and STL-10 [24], as well as two datasets comprising medical images, namely the Pneumonia Dataset [25], depicting pediatric chest X-ray images of patients with and without pneumonia, and the Melanoma Skin Cancer Dataset of the 2018 ISIC challenge [26, 27], depicting benign and different malignant melanoma images. Further details are described in Table 1. To allow a fair comparison with Nakata et al.'s method, we employ the same $k$ ($k = 10$) for the $k$-NN classifier throughout all our experiments.

[1] `vit_large_patch14_dinov2.lvd142m`
[2] `vit_base_patch14_dinov2.lvd142m`
[3] `https://huggingface.co/timm`

**Table 1**: Details of the selected datasets ($^*$ the reported resolution represents the average resolution across all samples).

| Dataset | # C | Resolution | # Train / Test |
|---|---|---|---|
| CIFAR-10 | 10 | $3 \times 32 \times 32$ | $50,000$ / $10,000$ |
| CIFAR-100 | 100 | $3 \times 32 \times 32$ | $50,000$ / $10,000$ |
| STL-10 | 10 | $3 \times 96 \times 96$ | $5,0000$ / $8,000$ |
| Pneumonia$^*$ | 2 | $1 \times 1328 \times 971$ | $5,232$ / $624$ |
| Melanoma | 7 | $3 \times 600 \times 450$ | $10,015$ / $1,513$ |

**Table 2**: Classification accuracy of our $k$-NN approach for different backbone choices.

| Accuracy [%] | CIFAR-10 | CIFAR-100 | STL-10 |
|---|---|---|---|
| ResNet-101 | 87.3 | 63.6 | 98.1 |
| CLIP ViT-B/16 | 92.4 | 68.0 | 98.5 |
| CLIP ViT-L/14 | 95.5 | 74.2 | 99.4 |
| DINOv2 ViT-B/14 | 98.0 | 87.2 | 99.4 |
| DINOv2 ViT-L/14 | **98.5** | **88.3** | **99.5** |

### 4.1. Backbone choice for our method

To validate our backbone choice, we compare the classification performance of our method with DINOv2 Large, to DINOv2 Base, a WideResNet101 [28] pre-trained on ImageNet-1k [29] as well as our own implementation of Nakata et al.'s ViT-B/16[3][4] and ViT-L/14[3][5] image encoder models (both pre-trained by CLIP). The results on CIFAR-10, CIFAR-100, and STL-10 are presented in Table 2. The results demonstrate the overall increased representative ability of models pre-trained in a self-supervised fashion compared to supervised pretraining. Moreover, both DINOv2 models, in particular DINOv2 ViT-L/14, showcase superior classification prowess compared to CLIP, endorsing the benefits of embracing self-supervised, image-exclusive pre-training for image-specific tasks.

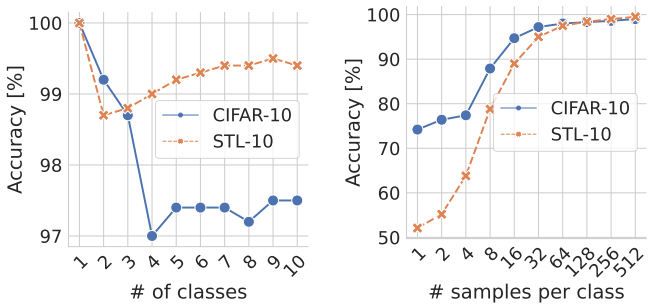
### 4.2. Adaptation for medical image analysis

To evaluate the applicability of our $k$-NN method in the medical domain, we first assess its classification performance on the Pneumonia and Melanoma dataset. We compare the performance of our approach with state-of-the-art, fully supervised benchmarks trained end-to-end. For Pneumonia, we compare to CovXNet by Mahmud et al. [30] and for Melanoma to Cassidy et al.'s EfficientNetB0 model [31]. The results are displayed in Table 3. Despite the distinct, transferred behavior of this task, DINOv2 does not employ any medical knowledge during training, our method demonstrates high classification potential, even surpassing the supervised state of the art for the Melanoma dataset.

[4] `vit_base_patch16_clip_224.openai`
[5] `vit_large_patch14_clip_336.openai`

**Table 3**: Comparison of our approach's strong transfer learning ability for medical image analysis. († refers to fully supervised models, trained end-to-end.)

| Accuracy [%] | Pneumonia | Melanoma |
|---|---|---|
| CovXNet† [30] | **98.1** | — |
| EfficientNetB0† [31] | — | 62.1 |
| Ours (DINOv2 ViT-B/14) | 88.1 | 68.5 |
| Ours (DINOv2 ViT-L/14) | 89.9 | **69.8** |



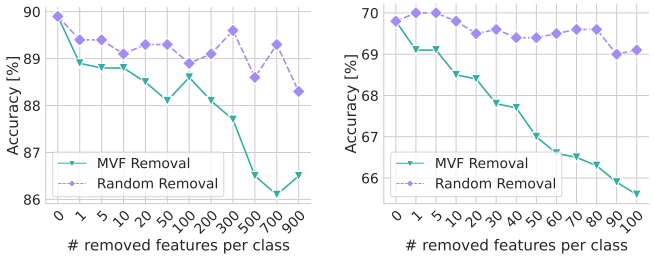

(a) Class incremental learning (b) Sample incremental learning

**Fig. 2**: Visualization of the method's ability for diverse continual learning tasks.

### 4.3. Continual learning and incremental forgetting

Nakata et al. [7] have shown that the $k$-NN approach promises the potential to mitigate catastrophic forgetting in continual learning scenarios for natural image datasets when incrementally adding additional classes or samples of existing classes to the support set. We first confirm this potential on CIFAR-10 and STL-10, by incrementally adding entirely new classes to the support and test set, as well as incrementally adding additional feature embeddings to the support set and evaluating the classification performance. Figure 2 (a) and Figure 2 (b) present the results for each task, respectively, showcasing the constant classification performance of our method for the class incremental learning task as well as a remarkable classification ability for the sample incremental learning task already for only a few samples per class in the support set. By using an adaptive $k$ instead of our fixed $k$, this few-shot classification capability could be improved even further.

Additionally, we evaluate the incremental learning capability of our method when transferring it to the medical domain. However, this time, we assess our method for incrementally adding datasets of different anatomies and distributions, instead of sticking to the same domain by adding additional classes of the same dataset. For this, we compare the method's exclusive performance on the Pneumonia and Melanoma dataset (cf. Table 3) with its performance on a combined version of both datasets, which comprises a diverse distribution in a multi-class classification setting. Notably, the




(a) Pneumonia (b) Melanoma

**Fig. 3**: Illustration of our method's classification consistency despite the continuous diminishing of the support set.

accuracy on the exclusive datasets is nearly consistent with the accuracy on the combined version (89.9 % vs. 89.9 % for Pneumonia and 69.8 % vs. 69.0 % for Melanoma).

Lastly, we investigate our approach's ability to facilitate the effortless removal of task-relevant data, ensuring minimal impact on the model's performance, particularly when patients exercise their rights to revoke data usage and demand the deletion of their information from the model. To this end, we evaluate the impact on classification performance if we remove either random samples or if we remove the most valuable feature embedding (MVF) of each class from the support set. One class's MVF is that feature embedding of the support set which the $k$-NN algorithm utilizes the most to correctly classify query samples during inference on a fixed test set. In other words, this feature embedding contributes the most to the classification performance of the method. Figure 3 visualizes this for the Pneumonia and the Melanoma dataset. The results present that removing nearly all support set samples poses only a slight, negative impact on the overall classification performance, demonstrating the few-shot ability of our model once again and thus demonstrating the overall potential of our method to remove any knowledge from our model without severely impairing the classification performance.

## 5. DISCUSSION AND CONCLUSION

In this work, we present an open-source, improved version of the $k$-NN integration with vision-based foundation models, originally proposed by Nakata et al. [7] that was not made publicly available by the authors before. Extensive experiments present our method's classification ability, apparent due to its high classification accuracy on natural images. We affirm its suitability for continuous learning scenarios, preventing catastrophic forgetting. Moreover, we showcase its potential for application in the medical domain, owing to its robust out-of-the-box performance and ability to seamlessly remove task-relevant data with minimal impact on performance. Our approach represents a significant step towards bridging the gap between foundation models' great performances and the challenges of data accessibility, privacy, and adaptability.

## 6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access. Ethical approval was not required as confirmed by the license attached with the open-access data.

## 7. REFERENCES


[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics*, 2019.

[2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D 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 Ziegler, Jeffrey Wu, Clemens Winter, Chris 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," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 1877–1901, Curran Associates, Inc.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[4] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever, "Learning transferable visual models from natural language supervision," in *International Conference on Machine Learning*, 2021.

[5] Mathilde Caron, Hugo Touvron, Ishan Misra, Herv'e J'egou, Julien Mairal, Piotr Bojanowski, and Armand Joulin, "Emerging properties in self-supervised vision transformers," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9630–9640, 2021.

[6] Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang, "Retrieval augmented language model pre-training," in *Proceedings of the 37th International Conference on Machine Learning*, Hal Daumé III and Aarti Singh, Eds. 13–18 Jul 2020, vol. 119 of *Proceedings of Machine Learning Research*, pp. 3929–3938, PMLR.

[7] Kengo Nakata, Youyang Ng, Daisuke Miyashita, Asuka Maki, Yu-Chieh Lin, and Jun Deguchi, "Revisiting a knn-based image classification system with high-capacity storage," in *Computer Vision – ECCV 2022*. 2022, pp. 457–474, Springer Nature Switzerland.

[8] Guowen Xu, Hongwei Li, Hao Ren, Kan Yang, and Robert H. Deng, "Data security issues in deep learning: Attacks, countermeasures, and opportunities," *IEEE Communications Magazine*, vol. 57, no. 11, pp. 116–122, 2019.

[9] European Union, "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation)," 2016, Article 17, Right to erasure ('right to be forgotten').

[10] Zhenyi Wang, Enneng Yang, Li Shen, and Heng Huang, "A comprehensive survey of forgetting in deep learning beyond continual learning," 2023.

[11] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "A continual learning survey: Defying forgetting in classification tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 07, pp. 3366–3385, 2022.

[12] Pádraig Cunningham and Sarah Jane Delany, "k-nearest neighbour classifiers - a tutorial," *ACM Computing Surveys (CSUR)*, vol. 54, pp. 1 – 25, 2020.

[13] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Neural Information Processing Systems*, 2017.

[14] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen, "Hierarchical text-conditional image generation with clip latents," *ArXiv*, vol. abs/2204.06125, 2022.

[15] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski, "Dinov2: Learning robust visual features without supervision," 2023.

[16] David Lopez-Paz and Marc'Aurelio Ranzato, "Gradient episodic memory for continual learning," in *Neural Information Processing Systems*, 2017.

[17] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy P. Lillicrap, and Greg Wayne, "Experience replay for continual learning," in *Neural Information Processing Systems*, 2018.

[18] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, G. Sperl, and Christoph H. Lampert, "icarl: Incremental classifier and representation learning," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5533–5542, 2016.

[19] Dara Bahri, Heinrich Jiang, and Maya Gupta, "Deep k-NN for noisy labels," in *Proceedings of the 37th International Conference on Machine Learning*, Hal Daumé III and Aarti Singh, Eds. 13–18 Jul 2020, vol. 119 of *Proceedings of Machine Learning Research*, pp. 540–550, PMLR.

[20] Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George van den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, Diego de Las Casas, Aurelia Guy, Jacob Menick, Roman Ring, T. W. Hennigan, Saffron Huang, Lorenzo Maggiore, Chris Jones, Albin Cassirer, Andy Brock, Michela Paganini, Geoffrey Irving, Oriol Vinyals, Simon Osindero, Karen Simonyan, Jack W. Rae, Erich Elsen, and L. Sifre, "Improving language models by retrieving from trillions of tokens," in *International Conference on Machine Learning*, 2021.

[21] Madhuri S. Mulekar and C. Scott Brown, *Distance and Similarity Measures*, pp. 385–400, Springer New York, New York, NY, 2014.

[22] Jeff Huber and Anton Troynikov, "Chroma - the open-source embedding database," 2023.

[23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, Eds. 2012, vol. 25, Curran Associates, Inc.

[24] Adam Coates, Andrew Ng, and Honglak Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Geoffrey Gordon, David Dunson, and Miroslav Dudík, Eds., Fort Lauderdale, FL, USA, 11–13 Apr 2011, vol. 15 of *Proceedings of Machine Learning Research*, pp. 215–223, PMLR.

[25] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y.L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A.N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, and Kang Zhang, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, 2018.

[26] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," 2019.

[27] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler, "Descriptor : The ham 10000 dataset , a large collection of multi-source dermatoscopic images of common pigmented skin lesions," 2018.

[28] Sergey Zagoruyko and Nikos Komodakis, "Wide residual networks," 2017.

[29] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, K. Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[30] Tanvir Mahmud, Md Awsafur Rahman, and Shaikh Anowarul Fattah, "Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization," *Computers in Biology and Medicine*, vol. 122, pp. 103869, 2020.

[31] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap, "Analysis of the isic image datasets: Usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, pp. 102305, 2022.