Energy and AI 20 (2025) 100492

Contents lists available at ScienceDirect

Energy and AI

journal homepage: www.elsevier.com/locate/egyai

Half-hourly electricity price prediction model with explainable-decomposition hybrid deep learning approach

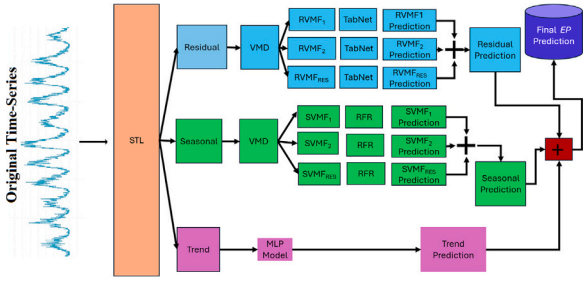
Sujan Ghimire [a], Ravinesh C. Deo [a], Konstantin Hopf [b], Hangyue Liu [c], David Casillas-Pérez [d,*], Andreas Helwig [e], Salvin S. Prasad [f,a], Jorge Pérez-Aracil [g], Prabal Datta Barua [a,h,i], Sancho Salcedo-Sanz [g,a]

[a] Artificial Intelligence Applications Laboratory, School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia
[b] University of Bamberg, Information Systems and Energy Efficient Systems, Kapuzinerstraße 16, 96047 Bamberg, Germany
[c] Sungrow Australia Pty Ltd, 1 Elizabeth Plaza, North Sydney, NSW 2060, Australia
[d] Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain
[e] School of Engineering, University of Southern Queensland, Toowoomba, QLD, 4350, Australia
[f] School of Sciences, College of Engineering, and Technical Vocational Education and Training, Fiji National University, Samabula, Suva, Fiji
[g] Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain
[h] Cogninet AI, Sydney, NSW 2010, Australia
[i] Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

HIGHLIGHTS

- Explainable artificial intelligence (*xAI*) half-hourly price prediction model is proposed.
- *xAI* model is optimized with the Optuna algorithm for enhanced accuracy.
- *xAI* model is benchmarked against fourteen stand-alone/decomposition models Deep analysis of results reaffirm excellent performance of xAI model.
- *xAI* has practical implications in electricity demand monitoring and forecasting systems.

GRAPHICAL ABSTRACT



ARTICLE INFO

Dataset link: https://www.aemo.com.au/

Keywords:
Tabular neural network
SHAP
Optuna algorithm
Deep learning
Machine learning
Convolutional neural network
LIME

ABSTRACT

Accurate prediction of electricity price (EP) is crucial for energy utilities and grid operators for enhancing the energy trading, grid stability studies, resource allocations and pricing strategies, thereby improving the overall grid reliability, efficiency, and cost-effectiveness. This study introduces a novel D3Net model for half-hourly EP prediction, integrating Seasonal-Trend decomposition using LOESS (STL) and Variational Mode Decomposition (VMD) with Multi-Layer Perceptron (MLP), Random Forest Regression (RFR), and Tabular Neural Network (TabNet). The methodology involves applying STL to the EP time-series to extract trend, seasonal, and residual components. The trend is predicted using an MLP model, the seasonal component is further decomposed with VMD into 20 Variational Mode Functions (VMFs) and predicted using an RFR model, and the residual component is decomposed with VMD and predicted using the TabNet model. Input features are identified using the Partial Autocorrelation Function , and models are optimized using the Optuna algorithm.

* Corresponding author.
E-mail addresses: sujan.ghimire@unisq.edu.au (S. Ghimire), ravinesh.deo@unisq.edu.au (R.C. Deo), konstantin.hopf@uni-bamberg.de (K. Hopf), henryliu0616@gmail.com (H. Liu), david.casillas@urjc.es (D. Casillas-Pérez), andreas.helwig@unisq.edu.au (A. Helwig), salvin.prasad@usq.edu.au (S.S. Prasad), jorge.perezaracil@uah.es (J. Pérez-Aracil), prabal.barua@unisq.edu.au (P.D. Barua), sancho.salcedo@uah.es (S. Salcedo-Sanz).

https://doi.org/10.1016/j.egyai.2025.100492
Received 5 November 2024; Received in revised form 27 January 2025; Accepted 3 March 2025
Available online 18 March 2025

The final prediction combines the trend, seasonal, and residual components' predictions. Explainable Artificial Intelligence (xAI) methods were used to enhance model interpretability and trustworthiness, with optimization via the Optuna algorithm. Comparative analysis with seven standalone and seven decomposition-based models confirmed the superior performance and statistical significance of the D3Net model. The D3Net achieved the highest global performance indicator for South Australia ($GPI \approx 11.068$) and Tasmania ($GPI \approx 12.206$). These results validate the efficacy and statistical significance of the D3Net model, demonstrating the viability of integrating STL and VMD decomposition approaches with MLP, RFR, and TabNet for EP prediction.

1. Introduction

With the deregulation of electricity markets, electricity prices (EP) have evolved into tradable commodities, necessitating accurate prediction for effective trading strategies and informed decision-making among market operators and participants. However, predicting EP is challenging due to the intricate dynamics of supply and demand, regional weather variations, and diverse market management strategies. These complexities lead to high price fluctuations, nonlinearity, and non-stationarity, posing significant hurdles for precise prediction. Improving prediction accuracy not only supports market stability but also aligns with Sustainable Development Goals (SDGs) such as affordable and clean energy (SDG 7) and industry, innovation, and infrastructure (SDG 9). Enhanced prediction capabilities facilitate better resource allocation, reduced energy wastage, and a more resilient energy infrastructure, thereby fostering sustainable development and economic growth.

Various methods are employed to predict EP, including traditional statistical methods, Machine Learning (ML) models, Deep Learning (DL) models, and hybrid approaches. EP is characterized by nonlinear historical data and the influence of large datasets on the accuracy of predictive algorithms. They exhibit weekly and daily seasonality, with variations during peak and off-peak hours. Nonetheless, predicting EP is still a complex task due to the ongoing demand for an uninterrupted power supply and the limitations in energy storage. According to [1], traditional statistical methods such as Autoregressive, Autoregressive-Moving-Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), seasonal ARIMA, Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and Vector Autoregressive Model (VAR) models are commonly employed for EP prediction. These models rely on optimized parameters, typically found using meta-heuristic algorithms like Particle Swarm Optimization [2]. Despite their simplicity and efficiency, these linear models struggle to capture the nonlinear relationships in EP, especially in high-frequency data such as hourly EP with significant fluctuations [3]. Consequently, multiple ML models have been proposed to confront these nonlinearities.

ML techniques address nonlinear relationships in EP by mapping inputs directly to outputs [4]. On this notion, several models, such as k nearest neighbor, Decision Tree (DT), Random Forest Regression (RFR), Support Vector Regression (SVR), and Artificial Neural Network (ANN) have been proposed for the domain of electricity price prediction. In [5], SVR was introduced as a model capable of nonlinear mapping in higher-dimensional spaces, offering efficient hourly electricity price forecasts. Similarly, [6] demonstrated that DT models effectively capture price spikes and forecast day-ahead EP in European markets. A study [7] highlighted RFR model, which, leveraging multiple decision trees, robustly handles outliers and noise, outperforming ARMA in real-time electricity price prediction. Furthermore, an extension of RFR, Extreme Gradient Boosting, proposed in [8], learns dependencies between current and historical data, surpassing ARIMA and SVR in hourly price prediction. The Work [9] evaluated the efficiency of ANN and SVR models using Australian National Electricity Market data, concluding that SVR trains faster and provides more accurate forecasts compared to ANN. In the Ontario electricity market, the study [10] found that k Least Squares Support Vector Machine based on SVR outperformed ANN, DT, and ARIMA models. Conversely, the configuration of ANN plays a critical role; studies [11,12] demonstrated that increasing the number of hidden layers significantly enhances EP prediction compared to Support Vector Machines, ARIMA, and VAR. Despite these ML models are known for their quick convergence and strong generalization capabilities, they frequently struggle in identifying time-related and deep-level hidden features. These features are crucial for capturing intricate temporal patterns and nuanced dependencies in EP data. For example, ANN can be vulnerable to overfitting without appropriate regularization. RFR tends to face increased computational complexity and difficulties in interpreting the ensemble of trees. DT are known for their susceptibility to high variance and instability with small changes in data. SVR can be sensitive to the choice of kernel parameters and may require careful tuning to achieve optimal performance, which adds to the computational complexity and tuning efforts involved in its application to EP prediction. These limitations highlight the ongoing challenges in effectively modeling and predicting the intricate dynamics of EP using ML models.

With the rising complexity of EP data and advancements in computer hardware, developing DL prediction models holds considerable promise due to their superior capabilities in extracting deep-level features and fitting nonlinear patterns compared to traditional ML methods [13]. DL models, a subset of ANN, typically include recurrent layers such as Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) [14]. RNNs leverage memory to process sequential data, making them effective for modeling dynamic EP fluctuations. However, RNNs can face challenges like gradient vanishing and explosion with longer sequences [15]. GRU and LSTM were introduced as solutions to these issues [16], with GRU often converging faster due to fewer parameters, though both generally achieve similar accuracy levels. These models are extensively utilized in EP prediction, with GRU showing slight advantages over LSTM in certain markets like Turkey and Europe [17], while LSTM, enhanced with techniques like signal decomposition and Bayesian optimization, outperforms other models in electricity markets such as Pennsylvania-New Jersey-Maryland interconnection and Denmark [18]. Despite their strengths, LSTM and GRU models face competition from statistical models like linear regression and GARCH, which have demonstrated superior accuracy in specific contexts such as the New Zealand market [19]. Additionally, Convolutional Neural Network (CNN) offer an alternative approach for modeling sequential data by extracting features through convolution and pooling operations [20]. While CNN alone may not surpass RNN-based models, integrating CNN with other techniques, such as dilated convolutional layers, has demonstrated potential in enhancing prediction accuracy, particularly in markets like Ontario [21].

While ML and DL models excel in handling nonlinear data compared to statistical models, they are prone to issues such as overfitting, local optima, and sensitivity to hyperparameters [22]. Furthermore, the inherent volatility of EP data poses a significant challenge for standalone ML and DL models. To mitigate these challenges, hybrid models have emerged to enhance prediction accuracy by combining diverse models or integrating optimization algorithm. Recent advancements in DL and hybrid models are summarized in Table 1. For instance, CNN-LSTM hybrid models have been investigated for predicting daily EP, demonstrating improved forecasts particularly for extreme values [23]. Similarly, the sequence-to-sequence (Seq2Seq) model based on LSTM layers

Table 1
Summary of recent models for EP prediction.

Ref.	Model	Inputs	Prediction horizon	Datasets
[28]	Multi-head self-attention and Convolutional Neural networks	168 hourly lagged EP values	Day ahead	Ontario electricity prices
[29]	LASSO Estimated AutoRegressive (LEAR)	Preceding 8 h of EP prices	Day ahead	Market for Ireland and Northern Ireland
[30]	(CNN-GRU) with an attention mechanism	Cyclical time variables and other climatic variables	Hourly	German electricity market
[27]	VMD-CLSTM-VMD-ERCRF	Record of EP of the past 24 h	Half-hourly	Queensland, Australia
[31]	CNN-GRU	Lagged EP, electricity load, generation, import and export, and weather data	Hourly	Ontario electricity prices
[32]	Quantile Regression and Kalman Filter	Lagged EP data and other climatic variables	Daily	Iranian wholesale market
[33]	Combined probability forecasting system (CPFS)	Lagged price data	Half-hourly	New South Wales, Australia and Singapore half-hourly electricity price dataset
[34]	CNN-LSTM	Record of EP of the past 24 h	Short-term (Hour ahead)	PJM Regulation Zone Preliminary Billing Data
[35]	CNN-LSTM	24, 168, and 720 hourly lagged EP values	Day, week, and month ahead	German electricity spot price
[23]	CNN-LSTM	168 hourly lagged EP values	Day ahead	Iranian electricity market
[36]	WT-Adam-LSTM	168 hourly lagged EP values	Hour ahead, Day ahead	New South Wales of Australia and French
[37]	WT-SAE-LSTM	Record of EP of the past 24 h	Short-term	U.S. Energy Information Administration (EIA)
[38]	SEPNet	Record of EP of the past 24 h	Hour ahead	New York City in the United States
[39]	ARD-ETR	168 hourly lagged EP values	Short-term	Nord Pool market
[40]	RVMs-LR	24, 168, hourly lagged EP values	Day ahead	New England Electricity market
[41]	LSTM	24, 168, hourly lagged EP values	Hour and Day ahead	PJM
[42]	LSTM	Record of EP for the past 96 h	Four hours ahead	Spanish electricity dataset
[43]	MD-Res.-EEMD -DE-ELM-DE-EL	Record of EP for the past 24, 48, 96, and 144 h	1, 2, 4, and 6 h ahead	Spanish and Australian electricity dataset
[44]	MOBBSA-ANFIS	168 hourly lagged EP values	Short-term	Ontario market

has been developed to capture the magnitude and timing of price spikes [24], with enhancements through attention mechanisms further improving prediction accuracy [25]. Transformer-based models have also shown effectiveness in daily EP forecasts, leveraging self and multi-head attentions to better capture long-term dependencies [26]. Additionally, hybrid approaches integrating mode decomposition with deep ML techniques have been successful in achieving high-accuracy forecasts, largely due to effective feature engineering [27].

Furthermore, researchers aiming to enhance EP prediction accuracy have explored additional strategies within hybrid models. These approaches include integrating advanced data preprocessing techniques to maximize model predictive power [45]. Signal decomposition methods, such as Wavelet Packet Decomposition [46], Empirical Mode Decomposition (EMD) [47], VMD [48,49], and STL [50], enable EP time series to be decomposed into seasonal, trend, and residual components. The STL decomposition method, specifically, offers robustness against outliers and effectively captures the seasonality, instability, randomness, and oscillations in EP, driven by meteorological factors [51]. By tailoring different models to each component extracted through STL decomposition, the precision of EP prediction results can be significantly enhanced.

Building on a comprehensive review of existing literature and model characteristics, this study introduces a novel ensemble system designed to enhance short-term EP prediction by optimizing data utilization and predictor potential. The proposed ensemble system consists of four main modules: data pre-processing, optimization, multi-scale prediction, and model interpretation. In the data pre-processing module, this study propose a hybrid decomposition approach that combines STL and VMD approaches. Given the seasonal, trend, and stochastic characteristics of the EP time series, STL first decomposes the series into three components: seasonal, trend, and residual. To further enhance predictability, VMD is then applied to the seasonal and residual components to break them down into Variational Mode Function (VMF) components. The multi-scale prediction module leverages three advanced ML models: MLP for the trend component, RFR for the seasonal component, and TabNet for the residual component. These models generate half-hourly EP predictions, which are then optimized using Optuna for hyperparameter tuning within the optimization module. The final EP prediction is derived by aggregating the outputs from each sub-model.

Addressing the challenge of model transparency, this study incorporates eXplainable Artificial Intelligence (*xAI*) techniques to enhance the interpretability and trustworthiness of the predictions [52]. Specifically, SHAP [53] are used for global interpretation, while LIME [54] provide local insights. These tools help elucidate the influence of individual features on half-hourly EP predictions. To assess the performance of the proposed model, referred to as D3Net (Decomposition and 3 models; MLP, RFR and TabNet), this study utilized EP time-series data from TAS and SA, comparing its predictions against fourteen other models through rigorous deterministic and statistical metrics analysis. This explainable decomposition-based model represents a robust framework for EP prediction, leveraging both advanced decomposition and ML techniques for enhanced accuracy and interpretability. The main contributions are as follows:

1. A novel ensemble system has been developed for half-hourly electricity price (EP) prediction. This system employs a hybrid decomposition approach to effectively extract the intrinsic characteristics of EP time-series data. By leveraging Optuna hyperparameter tuning approach and incorporating three advanced ML models, the system achieves stable and accurate predictions.
2. A novel hybrid decomposition approach, combining STL and VMD, is proposed for EP prediction. Given the strong seasonality and trends in EP influenced by various meteorological factors, STL is employed to isolate the seasonal, trend, and residual components. Following this, VMD is applied to further decompose the seasonal and residual components into a series of VMFs. The seasonal and residual series obtained from the STL algorithm represent the portion of the data remaining after extracting the trend, capturing irregular and unpredictable fluctuations. These components significantly influence the overall prediction results when combined.
3. Advanced ML models are employed as sub-models in this study, differing from previous research that typically used basic ML or simple linear statistical models. This study leverages advanced DL models (MLP and Tabular Neural Network) to enhance feature extraction and temporal modeling capabilities. Additionally, unlike many ensemble systems that require manual hyperparameter tuning, this study uses the Optuna optimization algorithm to optimize all sub-models, ensuring the stability and accuracy of the ensemble system.
4. The D3Net model's effectiveness and accuracy are validated by comparing its results with those of other decomposition-based and standalone ML and DL models.
5. Conduct a thorough post-hoc analysis using SHAP and LIME to enhance transparency and interpretability of the proposed D3Net model.
6. To aid the energy experts and grid operators with more accurate and interpretable predictions of EP that can help enhance the overall grid reliability, efficiency, and cost-effectiveness.

Before we describe the methodology in Section 3 and show the prediction results and model explanation using *xAI* in Section 4, the next section discusses the theoretical background.

2. Theoretical background

In this section, a brief overview of the decomposition algorithm and the model used to construct the proposed Decomposition and hybrid deep learning model for half-hourly EP prediction is presented. To rigorously benchmark the objective model, several decomposition-based ML models are utilized.

2.1. Seasonal-Trend decomposition using LOESS (stl)

The STL decomposition technique is a versatile and robust method used for decomposing time-series data. It partitions the time-series (Y_t) into three components: a residual component R_t with irregular variations, a low-frequency trend component T_t, and a high-frequency seasonal component S_t, see Eq. (1) [55]:

$$Y_t = S_t + T_t + R_t \tag{1}$$

STL aims to mitigate the effects of seasonal and long-term trends in the data, leaving behind segments that may contain potential anomalies or irregularities [56]. The pseudocode of the STL decomposition is shown in Algorithm 1. The calculation of the trend part T_t in the STL decomposition is calculated using the local weighted regression (Loess) method. The calculation formula is:

$$T(t) = \beta_0 + \beta_1 t \tag{2}$$

$$\beta_0 = \overline{Y} - \beta_1 \bar{t} \tag{3}$$

$$\beta_1 = \sum \frac{Y(i) - \overline{Y}}{t(i) - \bar{t}} \tag{4}$$

where $Y(i)$ is the observed value of the i-time point in the data set, $\bar{t}$ is the average of all time points in the data set, $\overline{Y}$ is the average of all observations in the data set and the seasonal part S_t in STL decomposition is calculated by moving average method. The calculation formula is as follows:

$$S(t) = \frac{[Y(t-1) + Y(t) + \cdots + Y(t-m+1) + Y(t-m)]}{m} \tag{5}$$

where m is the seasonal period length and $Y(t)$ is the observed value at time t.

Algorithm 1 STL Decomposition

Input: Time series $Y = \{Y_1, Y_2, ..., Y_T\}$, number of iterations K
Output: Trend $T[K]$, Seasonal $S[K]$, Deseasonalized $Y_{\text{deseasonalized}}$
Initialize $T, S \leftarrow$ empty arrays
for $k = 1$ to K **do**
  // Step 1: Detrending
  for each t in N **do**
    $Y_{\text{detrended}}[t] \leftarrow Y[t] - T[k-1][t]$ ▷ Subtract trend component from previous iteration
  end for
  // Step 2: Cycle-subseries smoothing
  $S_{\text{smoothed}} \leftarrow \text{LOESS\_smooth}(Y_{\text{detrended}})$
  // Step 3: Low-pass filtering smoothed cycle-subseries
  $S_{\text{moving\_avg}} \leftarrow \text{Moving\_Average}(S_{\text{smoothed}})$
  $L_{\text{trend}} \leftarrow \text{LOESS\_smooth}(S_{\text{moving\_avg}})$
  // Step 4: Detrending of smoothed cycle-subseries
  $S_{\text{updated}} \leftarrow S_{\text{smoothed}} - L_{\text{trend}}$
  // Step 5: Deseasonalizing
  for each t in N **do**
    $Y_{\text{deseasonalized}}[t] \leftarrow Y[t] - S_{\text{updated}}[t]$
  end for
  // Step 6: Trend smoothing
  $T[k] \leftarrow \text{LOESS\_smooth}(Y_{\text{deseasonalized}})$
end for
Return $T[K], S[K], Y_{\text{deseasonalized}}$

2.2. Variational Mode Decomposition (VMD)

VMD is a signal processing technique that decomposes signals into modes with auto-tuning capabilities. Unlike recursive methods, VMD is a fully non-recursive variational model where modes are extracted

simultaneously [57]. This approach effectively reduces the complexity of non-smooth and non-linear time series, breaking the data into multiple components with different frequencies [58]. VMD is more robust to sampling and noise compared to methods like EMD and its variants. The VMD method decomposes the original signal into distinct modes, known as VMFs, by minimizing the sum of the bandwidths of each mode. The decomposition process is framed within a variational framework to derive the targeted VMFs [59]. The optimal solution is achieved through a non-iterative sifting structure, which transforms the VMD procedure into solving a specific constrained variational problem. The problem to be addressed is as follows [27]:

$$\min_{\{v_k\},\{\omega_k\}} \{ \sum_{k=1}^{K} \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \|_2^2 \} \quad s.t. \sum_k v_k = s(t) \tag{6}$$

Here, the objective is to minimize the sum of the bandwidths of each mode, subject to the constraint that the sum of the modes equals the original signal $s(t)$. This formulation represents the VMD process, where $\{v_k\}$ are the modes i.e. VMFs, $\{\omega_k\}$ are the associated center frequencies, and $\delta(t)$ is the Dirac delta function. The convolution operation is denoted by $*$, and the notation $\| \cdot \|_2$ represents the L_2 norm, which is used to measure the bandwidth. The exponential term $e^{-j\omega_k t}$ represents the complex exponential function used in the decomposition.

The problem is further refined by introducing the augmented Lagrange function, transforming the constrained issue into a new form, described as follows:

$$L(\{v_k\}, \{\omega_k\}, \lambda) = \alpha \| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * v_k(t) \right] e^{-j\omega_k t} \|_2^2 + \| s(t) - \sum_{k=1}^{K} v_k(t) \|_2^2 + \langle \lambda(t), s(t) - \sum_{k=1}^{K} v_k(t) \rangle \tag{7}$$

where $\lambda(t)$ is the Lagrange multiplier used to enforce the constraint that the sum of the modes equals the original signal, α is a balancing parameter that controls the trade-off between the sum of the mode bandwidths and the fidelity to the original signal, and j is the imaginary unit, representing the square root of -1.

Eq. (7)is solved by the Alternating Direction Method of Multiplier. In this algorithm, the decomposed modes and center frequencies are transferred into the Fourier domain, which can be expressed as follows:

$$v_k^{(n+1)}(\omega) = \frac{s(\omega) - \sum_{i \neq k} v_i^{(n)}(\omega) + \frac{\lambda^{(n)}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^{(n)})^2} \tag{8}$$

$$\omega_k^{(n+1)} = \frac{\int_0^\infty \omega \left| v_k^{(n+1)}(\omega) \right|^2 d\omega}{\int_0^\infty \left| v_k^{(n+1)}(\omega) \right|^2 d\omega} \tag{9}$$

The updating of modes and center frequencies proceed until the condition in Eq. (10) is satisfied.

$$\frac{\sum_{k=1}^{K} \| v_k^{(n+1)} - v_k^{(n)} \|_2^2}{\| v_k^{(n)} \|_2^2} \leq \varepsilon \tag{10}$$

where ε is a small positive parameter, typically representing a tolerance or error threshold.

2.3. Random Forest Regression (RFR)

The RFR model, a machine learning algorithm based on DTs, is highly effective for regression tasks and variable impact assessment [60]. Central to this model is the Classification and Regression Trees algorithm, which operates under the principles of supervised learning. In an RFR, each DT independently selects features to partition node attributes, collectively forming a robust ensemble [61]. The model's prediction is derived from aggregating the outputs of all DT. Specifically, in regression tasks, the final prediction is the average of these outputs, expressed as:

$$\overline{H}(x) = \frac{1}{k} \sum_{i=1}^{k} \left[h_i(x = Y) \right] \tag{11}$$

where $\overline{H}(x)$ refers to a composite Random Forest model that aggregates predictions from multiple DT $h_i(x)$. Each $h_i(x)$ represents an individual DT within the ensemble. The variable Y represents the output or target variable of interest, and k denotes the total number of DT in the RFR model. The RFR algorithm is advantageous for addressing regression problems for several key reasons. Firstly, it demonstrates robust generalization capabilities by leveraging the collective wisdom of multiple trees, which helps guard against overfitting, rather than relying on a single model [62]. Secondly, unlike many traditional regression approaches, RFR do not require a predefined functional form, allowing them to more accurately capture complex, nonlinear relationships in the data. Moreover, these models are computationally efficient and can take advantage of parallel computing, facilitating faster training and prediction times, which is particularly beneficial for large datasets [63]. Additionally, RFR model provides insights into feature importance, helping to identify which variables have the most significant impact on the prediction outcomes. These characteristics make random forests a versatile and powerful tool for regression tasks in various domains.

2.4. Multi-Layer Perceptron (MLP) networks

The MLP is renowned for its versatility in addressing a broad spectrum of problems within neural network models. Functioning as a feed-forward neural network, the MLP is structured with one or more hidden layers, and their configurations can be adapted to suit the complexity of the training data [64]. The MLP architecture encompasses an input layer, an output layer, and intermediate hidden layers where the network operates by applying activation functions to weighted sums of input variables. Nodes within each layer are interconnected with all nodes in the subsequent layer through specific weights [65]. Learning within the perceptron involves adjusting these connection weights based on the discrepancy between the output and the expected result after processing the data [66]. This supervised learning process, known as backpropagation, extends the linear perceptron's least mean squares algorithm. The output y_k of the kth neuron in the output layer is expressed as:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F_j \left(\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1 \right) \quad \text{for } 1 \leq k \leq m \tag{12}$$

where w_{ij}^1 denotes the weights that connect the input and the hidden layers; b_j^1 and v_i^0 represents the threshold in hidden nodes and input supplied to the network; w_{jk}^2 denotes the weights that connect the hidden and output layer; n_i and n_h are the number of input nodes and hidden nodes; m represents the number of output nodes. F denotes the activation function.

2.5. Tabular neural network

The Tabular Neural Network (TabNet) model [67] represents an advanced deep learning architecture tailored for tabular data, renowned for its high accuracy and interpretability (Figs. 1a). It employs an end-to-end approach to streamline data processing while integrating sequential attention mechanisms for feature selection. This sequential attention provides clear insights into the rationale behind feature selection at each step, thus enhancing interpretability. Feature selection is driven by the attentive transformer layer, which generates a sparse and non-repetitive Mask matrix based on the outcomes of the previous steps. This variability in mask vectors across samples allows TabNet to accommodate diverse feature selections for different samples [68]. In the TabNet framework, the feature transformer layer manages the computation and processing of selected features from the current step. Unlike traditional DT structures that rely on simple feature thresholds, TabNet utilizes a more sophisticated feature transformer layer

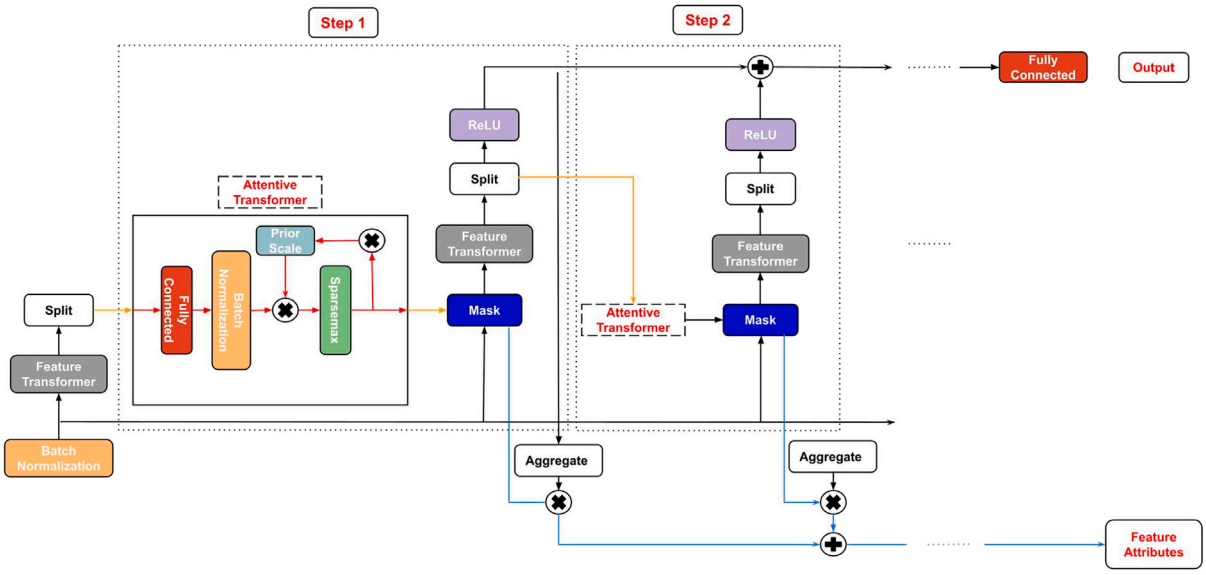


Fig. 1a. A schematic of TabNet Model.

to efficiently handle feature combinations, thereby surpassing decision trees in computational efficiency and feature integration capabilities. The design of TabNet involves a sequence of multiple decision steps, denoted as N_{steps}. In the $(i)^{th}$ step, the model uses the information processed in the $(i-1)^{th}$ step to determine which features to utilize. This step generates a processed feature representation that is then aggregated and employed for the overall decision-making process. The overall process is summarized as below [69]:

1. The dataset, characterized by a batch size B and D-dimensional features ($f \in \mathbb{R}^{B\times D}$), is provided as input to the model without any global feature normalization.
2. The input data first passes through a batch normalization layer and is subsequently delivered to a feature transformer.
3. The normalized features are processed using a feature transformer and subsequently split into two parts, $[\mathbf{d}[0], \mathbf{a}[0]]$, where $\mathbf{d}[0] \in \mathbb{R}^{B\times N_d}$ and $\mathbf{a}[0] \in \mathbb{R}^{B\times N_a}$.
4. Perform decision **Step 1**(Figs. 1a), $\mathbf{a}[0]$ serves as input to the attentive transformer, which produces the mask $\mathbf{M}[1] \in \mathbb{R}^{B\times D}$. This mask $\mathbf{M}[1]$ identifies the most important features, which are then processed through the feature transformer. The output from this processing is split into two parts: $\mathbf{d}[1]$, representing the decision step output, and $\mathbf{a}[1]$, providing information for the next decision step. Specifically, $\mathbf{d}[1] \in \mathbb{R}^{B\times N_d}$ and $\mathbf{a}[1] \in \mathbb{R}^{B\times N_a}$.
5. Similarly, for each subsequent decision step, the input is processed to yield the split features $[\mathbf{d}[i], \mathbf{a}[i]]$, where $\mathbf{d}[i] \in \mathbb{R}^{B\times N_d}$ represents the output for the current decision step, and $\mathbf{a}[i] \in \mathbb{R}^{B\times N_a}$ provides the necessary information for the following step.
6. Construct the overall decision embedding as

$$\mathbf{d}_{\text{out}} = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(\mathbf{d}[i]) \tag{13}$$

and apply a linear mapping $\mathbf{W}_{\text{final}}\mathbf{d}_{\text{out}}$ to obtain the final output.

7. The aggregate feature importance mask $\mathbf{M}_{\text{agg}-b,j}$ is obtained using the following formula:

$$\mathbf{M}_{\text{agg}-b,j} = \frac{\sum_{i=1}^{N_{\text{steps}}} \eta_b[i]\mathbf{M}_{b,j}[i]}{\sum_{j=1}^{D}\sum_{i=1}^{N_{\text{steps}}} \eta_b[i]\mathbf{M}_{b,j}[i]}, \tag{14}$$

where $\mathbf{M}_{b,j}[i]$ denotes the element located at the bth row and jth column of the mask in the ith decision step. The aggregate decision contribution $\eta_b[i]$ for the ith decision step of the bth sample is calculated as follows:

$$\eta_b[i] = \sum_{c=1}^{N_{dc}} \text{ReLU}(\mathbf{d}_{b,c}[i]), \tag{15}$$

where $\mathbf{d}_{b,c}[i]$ represents the element in the bth row and cth column of the split features for the ith decision step.

As aforementioned, TabNet model fundamental components comprise the Attentive Transformer layer and the Feature Transformer layer. The Attentive Transformer consists of four key layers: Fully Connected, Batch Normalization, prior scales, and Sparsemax. Initially, input passes through a Fully Connected layer post Batch Normalized. Subsequently, it is scaled by the prior scales, which aggregate relevant feature weights from preceding decision stages. The role of the Attentive Transformer is to compute the mask layer for the current step based on the outcomes of the preceding step. According to the structure of the Attentive Transformer shown in Figs. 1b, the feature selection mask is defined as:

$$\mathbf{M}[i = \text{sparsemax}(\mathbf{P}^{[i-1]} \cdot h_i(\mathbf{a}[i-1])) \tag{16}$$

where sparsemax denotes the sparsemax normalization function, $\mathbf{a}[i-1]$ represents the processed features from the preceding decision step, and h_i is a trainable function with a Fully Connected layer followed by Batch Normalization. The term $\mathbf{P}[i-1]$, referred to as the prior scale, indicates the historical usage of each feature up to step $i-1$, and is defined as:

$$\mathbf{P}(i-1) = \prod_{j=1}^{i-1}(\gamma - \mathbf{M}(j)) \tag{17}$$

where γ is a relaxation parameter. Initially, $\mathbf{P}^{(0)}$ is initialized as all ones: $\mathbf{P}[0] = \mathbf{1}^{B\times D}$. The Feature Transformer block (Figs. 1c) consists of several layers, some of which are shared across all decision steps while others are specific to each step. Each block includes a Batch Normalization layer, a Gated Linear Unit activation, and Fully Connected layers. The Gated Linear Unit is augmented with a normalization residual connection to stabilize the variance transformation in the network output, thereby preventing information loss and maintaining stability during the learning process. The Feature Transformer layer can be divided into two parts: the first half shares parameters across all steps, facilitating the extraction of common features. This design minimizes the need for parameter updates, thereby boosting the robustness of the

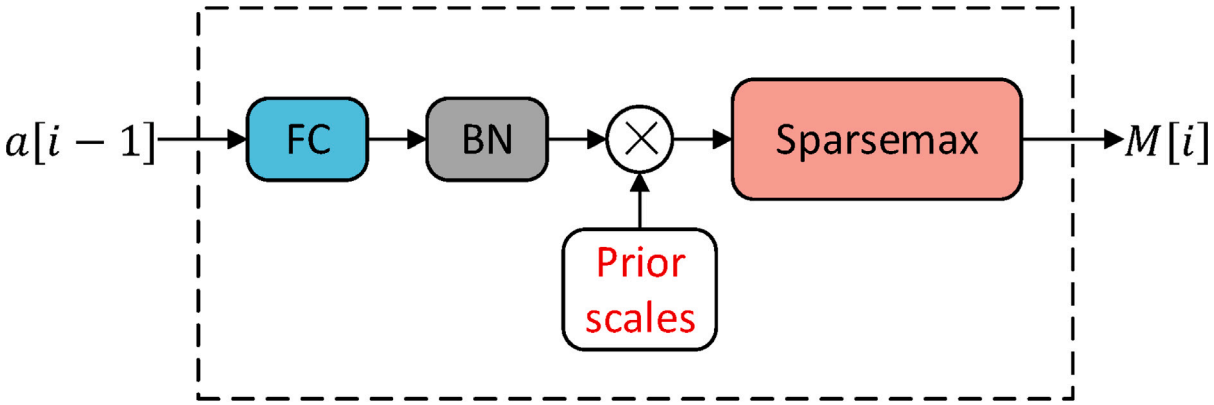


Fig. 1b. A schematic of Attentive Transformer in TabNet Model.

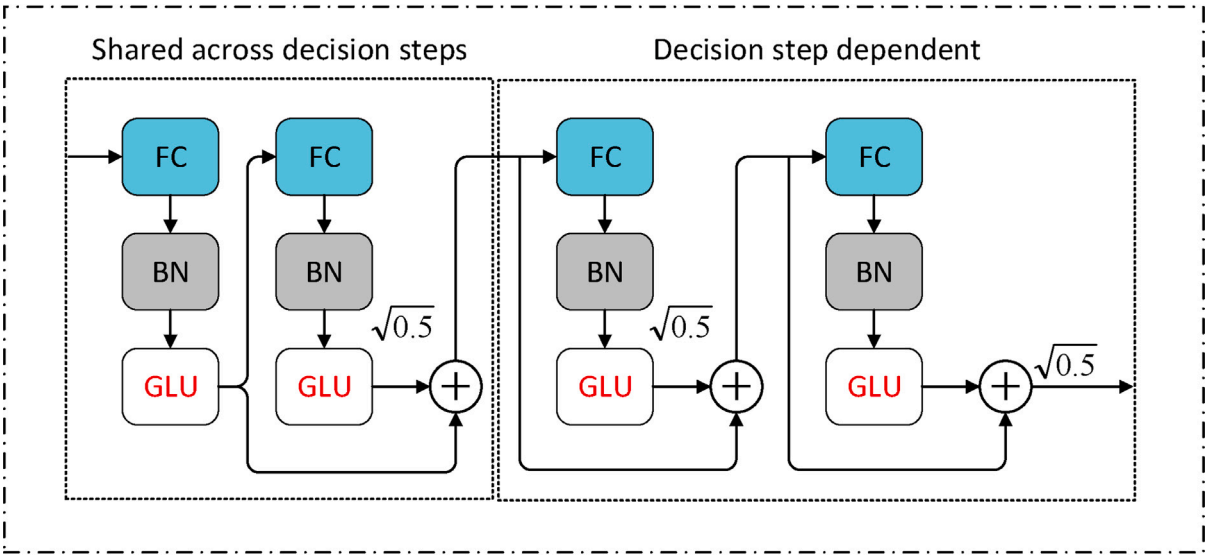


Fig. 1c. A schematic of Feature Transformer in TabNet Model.

learning process. The second half of the Feature Transformer layer is not shared and trained independently at each step. This allows for step-specific feature adjustments while initially leveraging shared layers to compute common feature aspects.

2.6. Objective model: Decomposition and hybrid deep learning network

The Decomposition and Hybrid Deep Learning Network, known as D3Net, employs STL decomposition initially on the EP series, followed by additional decomposition using VMD for the residual and seasonal components. In this approach, MLP is utilized for STL decomposed trend prediction, RFR for VMD decomposed seasonal prediction, and TableNet for VMD decomposed residual prediction, aiming to achieve robust and accurate EP time-series prediction through the synergistic capabilities of these models (Fig. 2). Further decomposition of the residual and seasonal components plays a critical role in improving prediction accuracy. STL decomposition often struggles to capture the intricate patterns or noise present in the residual component. By applying VMD, these complex patterns can be isolated, enabling more precise modeling and prediction. Similarly, the seasonal component may encompass multiple seasonal cycles or subtle patterns that require detailed separation for accurate predictions.

This iterative decomposition process not only enhances prediction accuracy but also improves data quality by reducing noise and clarifying the core signal. Matching each refined component with the most appropriate model, such as an ANN for nonlinear relationships, a Random Forest for complex interactions, or TableNet for residual patterns, maximizes the strengths of each model type. Furthermore, this method enhances the interpretability of time series data by revealing its underlying structure and distinguishing among different sources of variability. It transforms non-stationary components into more manageable, stationary forms, thereby facilitating more effective modeling and prediction. Ultimately, by integrating advanced decomposition techniques with tailored modeling approaches, D3Net ensures a sophisticated and reliable framework for achieving high-accuracy time-series prediction in academic and practical applications.

2.7. Benchmark models

1 Convolution Neural Network: A CNN is a specialized type of ANN designed primarily for processing grid-like data, such as images. It consists of multiple layers, including convolutional layers that apply filters to input data, pooling layers that downsample feature maps, and fully connected layers that perform classification or regression tasks [70,71]. In regression tasks, CNNs predict continuous outputs by learning hierarchical representations of features, extracting patterns relevant to the target variable from input data through convolution and pooling operations [72]. This architecture makes CNNs effective for tasks requiring spatial or sequential dependencies, such as time series prediction and image analysis [59,73].

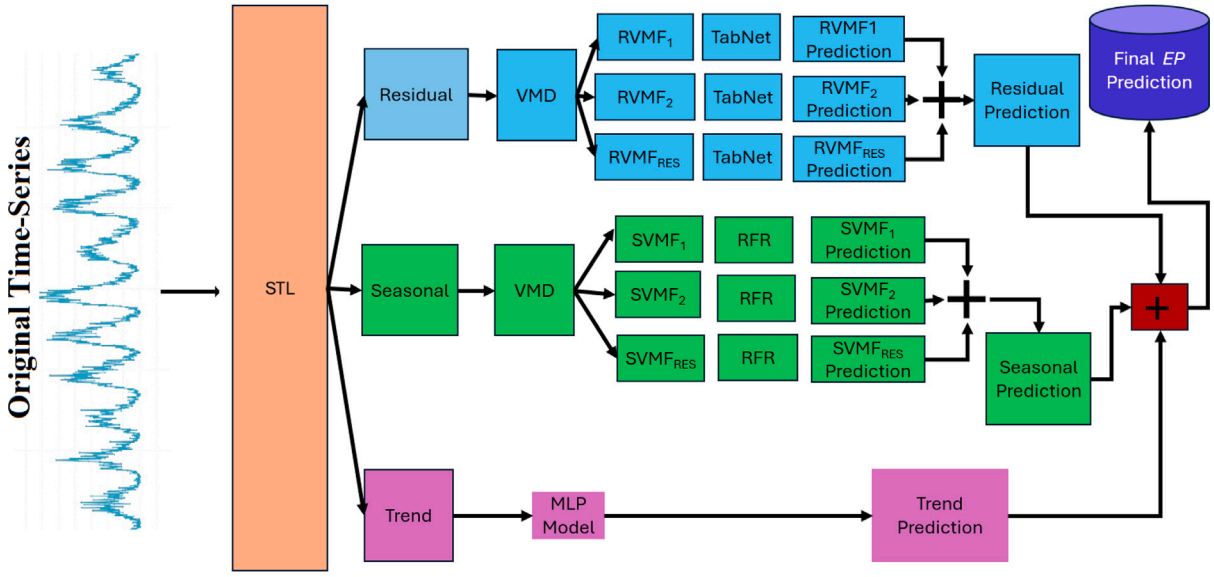


Fig. 2. Schematic structure of the Decomposition and hybrid deep learning network integrating Seasonal trend decomposition using Loess and VMD with MLP, Random Forest and TabNet model, where $\oplus$ is the aggregation.

2 LSTM is a type of RNN architecture designed to overcome the limitations of traditional RNNs in capturing long-term dependencies [74]. It includes specialized units called cells that maintain a memory state, enabling it to store and access information over long sequences. LSTMs use gates to control the flow of information into and out of the cells, which helps in learning and retaining patterns over time [27,75].

3 Deep Neural Network (DNN) is a type of ANN characterized by multiple layers between the input and output layers. Each layer typically consists of multiple neurons (nodes) that use activation functions to transform input data into meaningful representations for the subsequent layers [76,77]. DNNs can learn intricate patterns and hierarchical structures from data through a method known as DL. They perform exceptionally well in tasks such as image and speech recognition, natural language processing, and other domains where learning intricate features from large datasets is essential. Training deep neural networks often requires significant computational resources and large amounts of data to achieve high performance [78,79].

4 Decomposition based models: The trend, seasonal, and residual components decomposed via STL are predicted using several models: CNN, LSTM, DNN, RFR, MLP, and TabNet. These models are denoted as STL-CNN, STL-LSTM, STL-DNN, STL-RFR, STL-MLP, and STL-TabNet, respectively.

2.8. Optuna

Optuna is an open-source hyperparameter optimization framework implemented in Python, designed to automate the tuning of hyperparameters for ML and DL models. It offers a versatile, efficient, and user-friendly interface for defining and optimizing complex search spaces, making it especially suitable for models with numerous hyperparameters [80]. Optuna leverages advanced optimization techniques, such as Bayesian optimization and Tree-structured Parzen Estimator, to efficiently navigate the search space and identify optimal hyperparameter configurations. It also facilitates distributed optimization, pruning of non-promising trials, and compatibility with various ML libraries, boosting model performance and accelerating the development of reliable ML solutions. Detailed information on the optimization capabilities of Optuna can be found in [81]. The selection of Optuna for this study is based on its define-by-run context, efficient sampling, and ease of setup. The define-by-run context enables dynamic construction of the search space. During the optimization process, various hyperparameter combinations are evaluated to maximize a function and yield a validation score [82]. Optuna's efficient sampling handles both relational sampling, which benefits from parameter correlations, and independent sampling, which treats each sample separately [80]. In addition, Optuna's easy setup supports for a range of applications, from lightweight experiments through interactive interfaces to heavyweight distributed computing [83]. Both the objective model and benchmark models in this study are optimized using Optuna.

3. Methodology

This section provides a comprehensive overview of the methodologies employed in the development and evaluation of the D3Net model for predicting half-hourly EP. It begins with a detailed description of the data series used. Next, it outlines the various phases involved in designing the D3Net model. This is followed by an explanation of the robust statistical metrics used to assess the model's performance. Finally, it discusses the utilization of explainable Artificial Intelligence (xAI) model-agnostic tools to interpret the predictions made by the proposed D3Net model.

3.1. Electricity price data

The data utilized in this study were sourced from the Australian Energy Market Operator, providing comprehensive information on EP and total electrical consumption in Australia. In this study, data covering the years 2019 through 2022 was utilized to predict EP in two Australian states: Tasmania (TAS) and South Australia (SA). Summary statistics of the historical data are detailed in Table 2. Each dataset consistently contains approximately 17,520 entries per year, ensuring robust data coverage. Central tendency measures reveal significant year-over-year differences in mean and median values. In SA, the mean EP increased substantially from 47.21 Australian Dollar (AUD) per MWh in 2020 to 153.6 AUD/MWh in 2022, with the median following a similar upward trend. TAS showed a comparable pattern, with the mean rising from 42.86 AUD/MWh in 2020 to 137.36 AUD/MWh in 2022, though the absolute values were slightly lower. Variability in the data is captured by the standard deviation and interquartile range. SA's standard deviation surged from 37.43 AUD/MWh in 2020

Table 2
Summary statistics of the South Australia and Tasmania electricity prices (EP; AUD/MWh) of 2019–2021. ADF Statistic is the Augmented Dickey–Fuller (ADF) test statistic for unit root test and ADF p-value is the corresponding p-value at 95% confidence level.

Statistics	South Australia			Tasmania		
	2020	2021	2022	2020	2021	2022
Data points	17 568	17 520	17 520	17 568	17 520	17 520
Minimum	1.01	1	1.03	1.01	1.02	1
25% percentile	30.62	18.11	47.1	27.26	18.92	58.34
50% percentile	42.13	40.93	114.69	40.29	30.32	103.71
75% percentile	56.01	68.19	231.44	51.93	42.92	191.6
Mean	47.21	53.76	153.6	42.86	35.75	137.36
Maximum	431.01	981.65	998.77	878.4	675.47	995.2
Interquartile range	25.39	50.08	184.34	24.68	24	133.27
Standard deviation	37.43	55.17	140.14	28.21	30.91	111.57
Variance	1401.05	3043.83	19 638.77	795.95	955.69	12 447.14
Mean absolute deviation	21.03	35.96	108.55	17.97	18.68	85.73
Median absolute deviation	12.63	24.85	84.95	12.28	11.97	55.08
Skewness	3.99	3.5	1.5	4.49	4.55	1.58
Kurtosis	23.37	26.38	3.24	71.26	44.52	4.06
ADF Statistic	−9.92	−7.68	−5.19	−8.33	−8.04	−4.56
ADF p-value	0	0	0	0	0	0

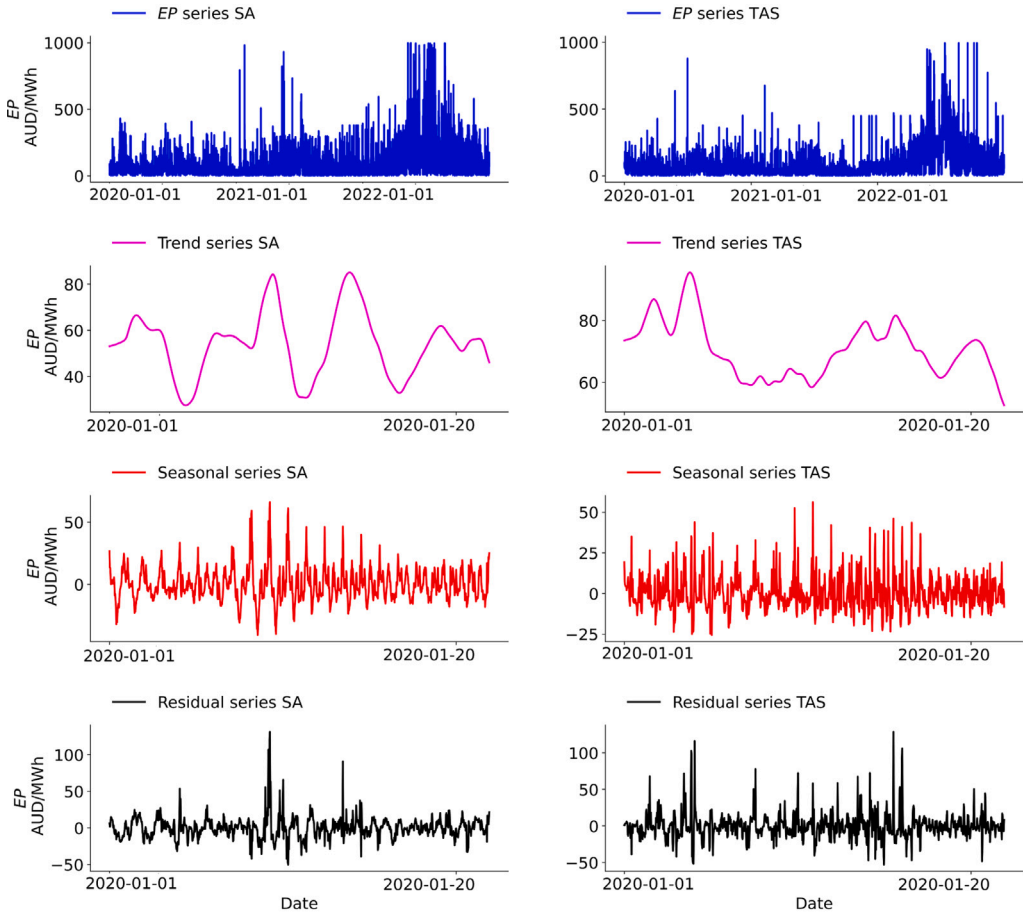


Fig. 3. Results from the application of the STL decomposition procedure to the EP time-series for SA and TAS. Top Panel: Original time series data for EP, Second Panel: Smoothed EP highlighting long-term trends without short-term fluctuations, Third Panel: Seasonal variations in the prices, indicating periodic patterns, and Bottom Panel: Residuals, which are the remaining variations after removing the trend and seasonal components, representing irregular or random fluctuations.

to 140.14 AUD/MWh in 2022, indicating greater data dispersion. The inter-quartile range also widened significantly, suggesting increased variability within the middle 50% of the data distribution. TAS exhibited a similar trend, with both standard deviation and inter-quartile range expanding over the same period. The distribution traits of the data are further highlighted by skewness and kurtosis statistic. Both datasets displayed high positive skewness, particularly in 2020 and 2021, indicating a right-skewed distribution with a long tail of high values. Kurtosis values indicated leptokurtic distributions characterized by heavy tails and sharp peaks around the mean. However, a decline in both skewness and kurtosis during 2022 suggests a shift towards a more normal distribution. Furthermore, the Augmented Dickey–Fuller (ADF) test produced highly negative statistics and p-values of zero across all datasets and years, strongly rejecting the null hypothesis of a unit root. This indicates that the datasets are stationary over the observed period, which is crucial for effective time series analysis and prediction.

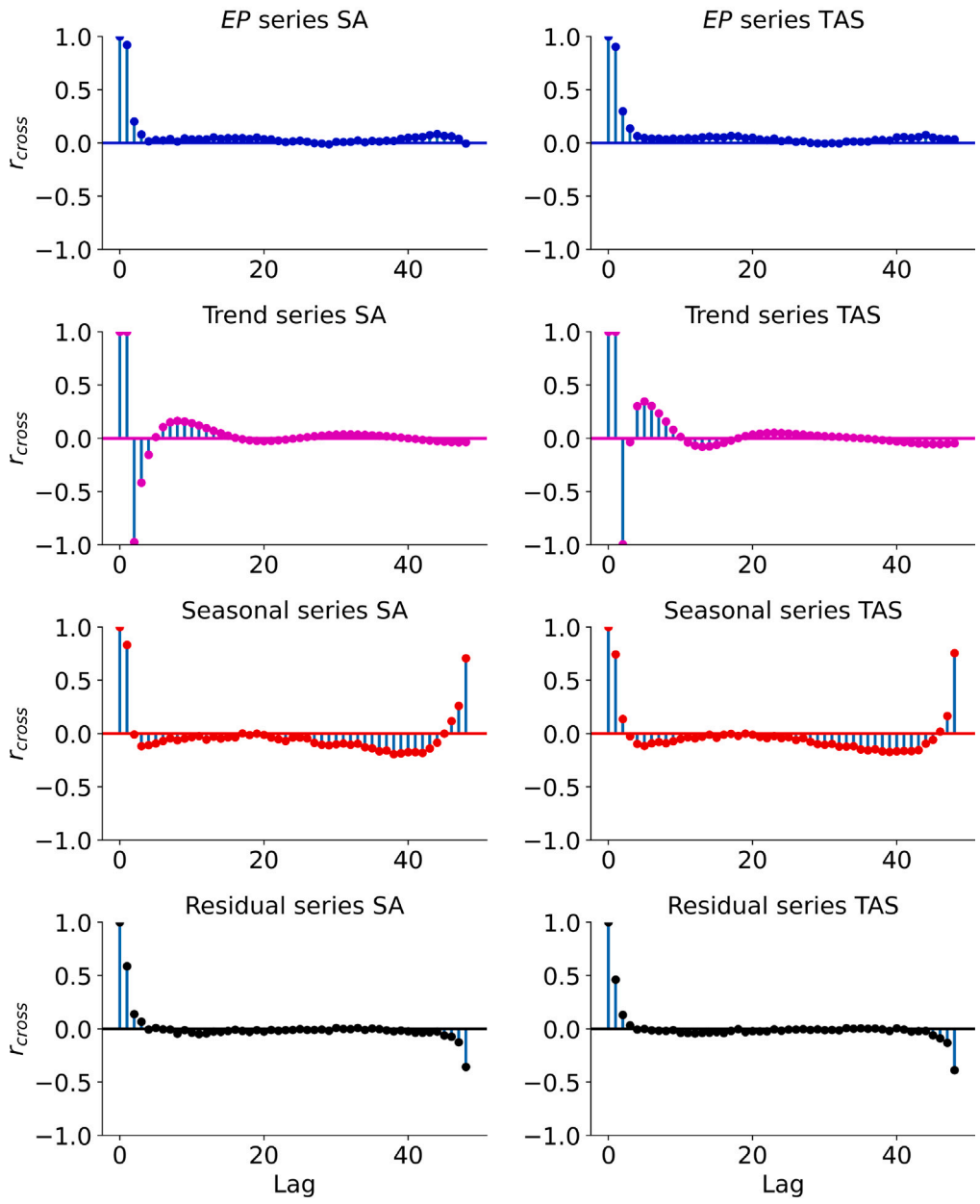


Fig. 4. PACF plot of STL decomposition series for SA and TAS. Trend Component: Persistent correlations indicating long-term dependencies, Seasonal Component: Periodic spikes at regular intervals, confirming the seasonal pattern and Residual Component: Largely uncorrelated, suggesting that the residuals are mostly random noise after removing the trend and seasonal components, with a few minor autocorrelations.

3.2. Data decomposition using STL

The EP time-series datasets are decomposed using the STL technique. This decomposition is illustrated in Figs. 3, which presents the original time-series data alongside the trend, seasonal, and residual components for both SA and TAS. The STL decomposition, as shown in Fig. 3, reveals that the trend component exhibits simpler and more predictable patterns compared to the seasonal and residual series. Notably, the residual component of the EP series displays a considerable range of oscillations, spanning from −50 to 100 AUD/MWh, while the seasonal series ranges from −25 to 50 AUD/MWh for both SA and TAS. This variability in the seasonal and residuals indicates more complex and irregular fluctuations in the EP after accounting for the trend effects.

Furthermore, the PACF plots for the trend, seasonal, and residual series after STL decomposition are depicted in Figs. 4. The PACF plot for the trend component displays significant values at the initial lags that gradually decline, indicating persistent correlations over several time periods (second panel in Figs. 4). This pattern suggests that the trend component is smoother and possesses longer-term dependencies, characteristic of trend data. The PACF plot for the seasonal component reveals significant spikes at specific lags corresponding to the seasonal cycle (third panel in Figs. 4). This reveals periodic behavior in the seasonal component, with correlations peaking at regular intervals, thereby confirming a seasonal pattern in the data. In contrast, the PACF plot for the residual component shows that most partial autocorrelations are close to zero, with a few exceptions (fourth panel in Figs. 4). With residuals being largely uncorrelated, it can be inferred that the remaining variations are mostly random noise after accounting for the effects of trend and seasonality. While there are significant spikes at a few lags, these are relatively minor compared to those observed in the trend and seasonal components.

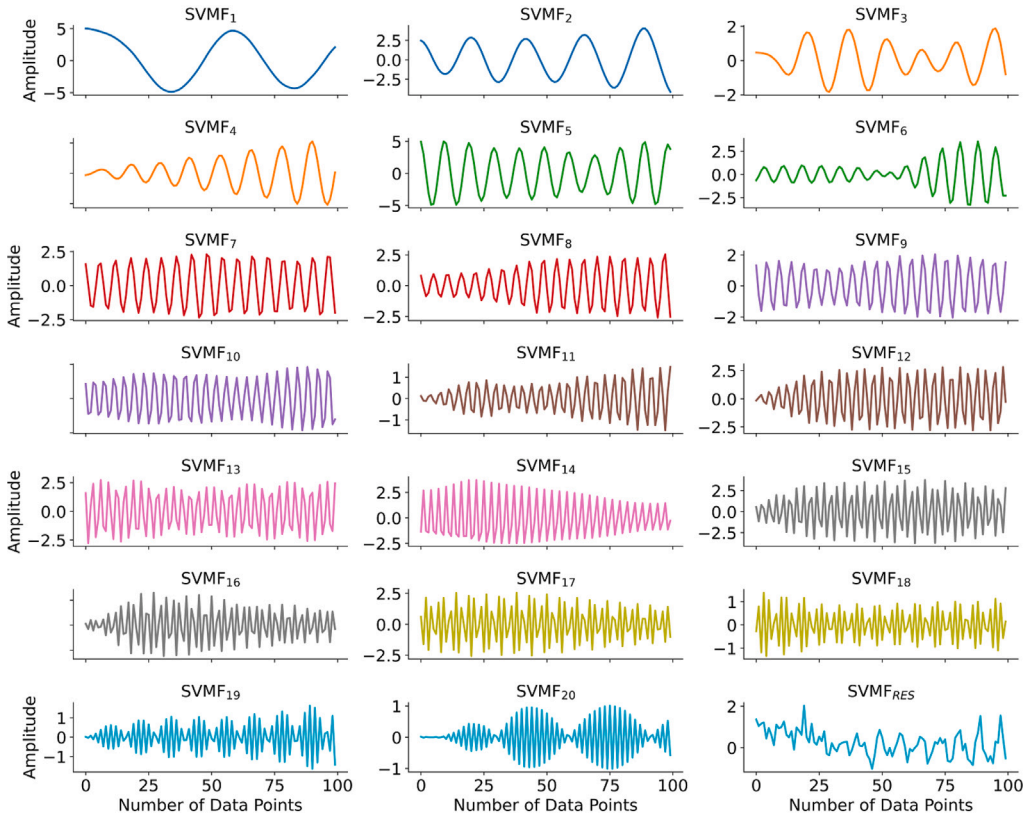


Fig. 5a. VMD decomposition results for seasonal component of STL for SA, where $SVMF$ is the seasonal variational mode function.

3.3. VMD decomposition of seasonal and residual series

As mentioned earlier, the VMD decomposition method is applied on the seasonal and residual elements of the EP series following its decomposition by STL. In VMD, the critical parameters include the Number of Modes (K), Penalty Factor (α), and Tolerance (τ). The parameter K determines the number of modes into which the signal will be decomposed. The penalty factor α regulates the bandwidth of each mode, with higher values leading to narrower bandwidth modes, ensuring that each mode effectively captures a specific frequency range. The tolerance τ is utilized as a convergence criterion for the algorithm. In this study, α is set to 10^3 and τ is set to 10^{-7}.

To determine the optimal number of modes, K is chosen based on the criterion of achieving the minimum Root Mean Square Error (RMSE). This approach involves varying K from 5 to 40 and evaluating the decomposition performance for each value by calculating the RMSE between the original signal and the reconstructed signal from the decomposed modes. The value of K that results in the lowest RMSE is considered optimal, as it indicates the best balance between capturing essential signal characteristics and avoiding overfitting or underfitting the data. The Fig. 5a illustrates the VMD of an EP seasonal series after STL for SA. The VMD process decomposes the time series into multiple VMF components, ranging from VMF_1 to VMF_{20}, with each one representing a distinct component of the decomposed seasonal series data. These VMFs display different frequency and amplitude characteristics, capturing various underlying patterns and oscillations within the decomposed seasonal series. The final plot, labeled VMF_{RES}, represents the residual component, highlighting the part of the series not captured by the previous VMFs. Each subplot shows the amplitude variations over the number of data points, allowing for a detailed examination of the decomposed signals' behaviors. Similarly, Fig. 5b shows the VMD of an EP residual series after STL for SA.

3.4. Data preprocessing

After the data decomposition, the decomposed data are normalized using the Z-score method. Z-score normalization, also known as standard score or Z-value, is a statistical technique that transforms the data into a standard normal distribution with a mean of zero and a standard deviation of one. The Z-score normalization formula is given by:

$$Z_i = \frac{X_i - \mu}{\sigma} \tag{18}$$

where Z_i is normalized score of the decomposed series, X_i is original value of the decomposed series, μ is a mean and σ is a standard deviation of the decomposed series.

3.5. Predictive model development

Once the dataset is normalized to standardize features and address scale discrepancies, the subsequent step involves partitioning the data into distinct subsets: training, validation, and test sets. Specifically, the training set encompasses the period from January 1, 2021, to December 23, 2023. This extensive time frame is crucial as it allows machine learning models to learn from historical patterns and relationships embedded within the data. To enhance model training and prevent overfitting, 20% of the training data, selected randomly, is designated as the validation set. This subset facilitates the fine-tuning of model parameters and performance evaluation during training. On the other hand, the test set comprises the final week of available data, spanning from December 24, 2023, to December 31, 2023. This segment of the dataset remains unseen during the training and validation stages, providing an unbiased evaluation of the model's predictive capabilities on new data.

After partitioning the dataset into training, validation, and test sets, the next step in this study involved a detailed analysis using correlogram plots to explore the PACF. This statistical tool helped us identify

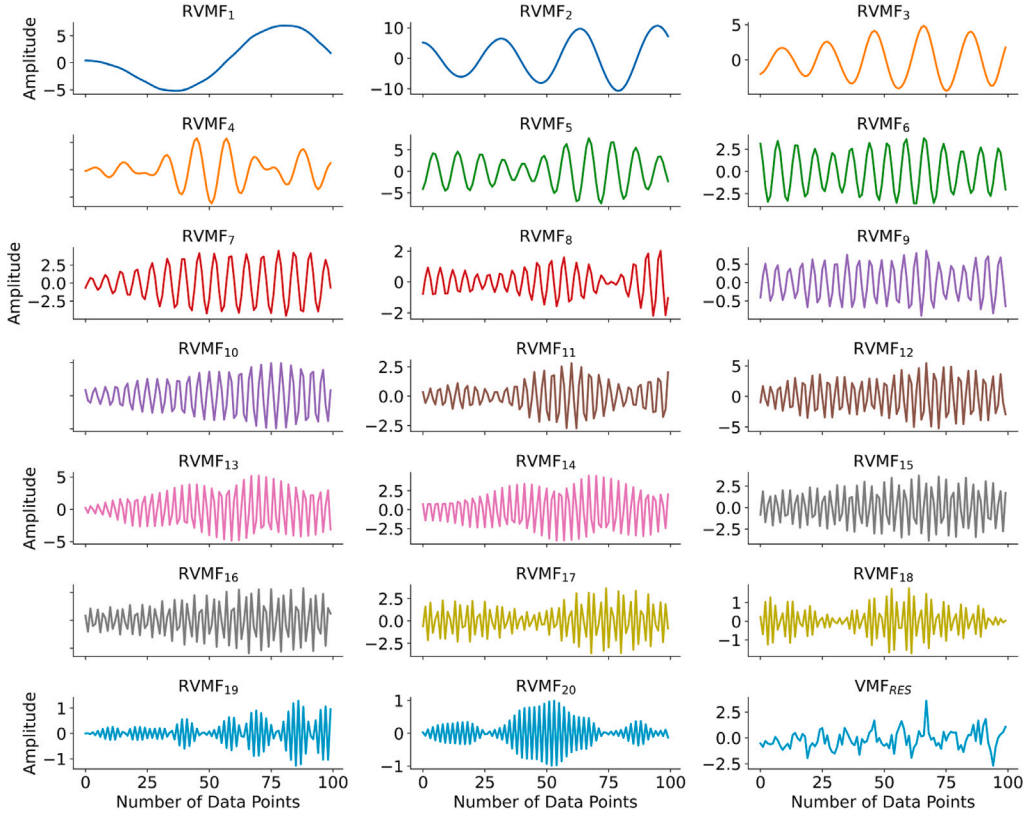


Fig. 5b. VMD decomposition results for residual component of STL for SA, where $RVMF$ is the residual variational mode function.

Table 3
Input parameter for the model development.

Model	Model name	Component	Model input	Target
Standalone Models	RFR, MLP, DNN, TabNet, LSTM, CNN, CLSTM		$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-10)}$	X_t
STL Decomposition Based Models	STL-CLSTM, STL-MLP, STL-RF, STL-DNN, STL-TabNet, STL-CNN, STL-LSTM	Trend	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-5)}$	X_t
		Seasonal	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-10)}$	X_t
		Residual	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-14)}$	X_t
D3Net	STL-VMD-MLP-RFR-TabNet	Trend	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-5)}$	X_t
		VMD Seasonal	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-10)}$	X_t
		VMD Residual	$X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-10)}$	X_t

and understand the significant time lags that influence half-hourly EP. In Figs. 6a and 6b, this study presented the PACF plot derived from decomposed seasonal and residual time-series data, specifically tailored to SA. The visualization offered valuable insights into the temporal dependencies and antecedent behaviors affecting EP, which are critical for designing precise prediction models.

Upon thorough examination of the PACF plots for the decomposed time-series, this study pinpointed the ten most consequential lagged series. These lagged series represent the historical patterns and previous values of EP that exhibit the strongest predictive power for future prices. By selecting these influential lagged series as model inputs, the most significant temporal relationships inherent in the data are captured and leveraged. This approach ensures the modeling strategy incorporates relevant historical information, thereby enhancing the accuracy and reliability of the prediction for half-hourly EP in SA and TAS. Table 3 presents the input parameter for each model.

3.6. Optuna hyperparameter optimization

Once the model inputs are identified, the first step in hyperparameter optimization is to define the objective function. The objective function includes the machine learning and deep learning logic of the program, which in this study encompasses the models under investigation (CNN, LSTM, Convolutional LSTM (CLSTM), TabNet). The objective function returns an evaluation score (RMSE), known as the objective value or the metric being optimized. The second step involves defining the model hyperparameter ranges. Optuna supports various data types for hyperparameters, including string, float, integer, and boolean. Each variable must have a specified range of values, which can be numerical or a list of strings. The hyperparameter ranges used in this study are provided in Table 4. After defining the objective function and hyperparameter ranges, an Optuna study systematically tests different combinations. The study is customizable, allowing the specification of the number of trials, the selection of an optimization algorithm (Tree-structured Parzen Estimator), and setting the optimization direction (minimize RMSE). This study also employs a pruning function to skip low-performing trials, saving time with large combinations.

3.7. Evaluation indexes

The study employed a comprehensive set of performance metrics to rigorously evaluate the newly constructed D3Net model against benchmark models in predicting half-hourly EP for the two Australian states (SA and TAS). These statistical metrics included the Coefficient of

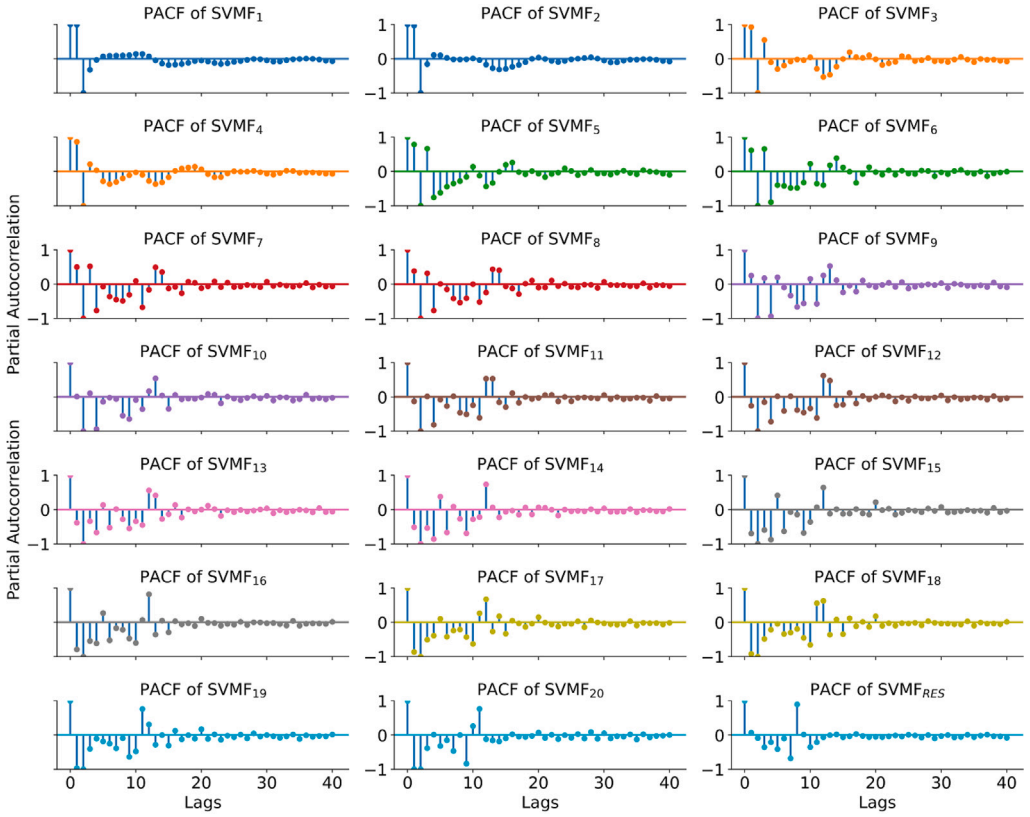


Fig. 6a. PACF of the decomposed seasonal series showing the most significant lags for the SA, where $SVMF$ is the seasonal variational mode function.

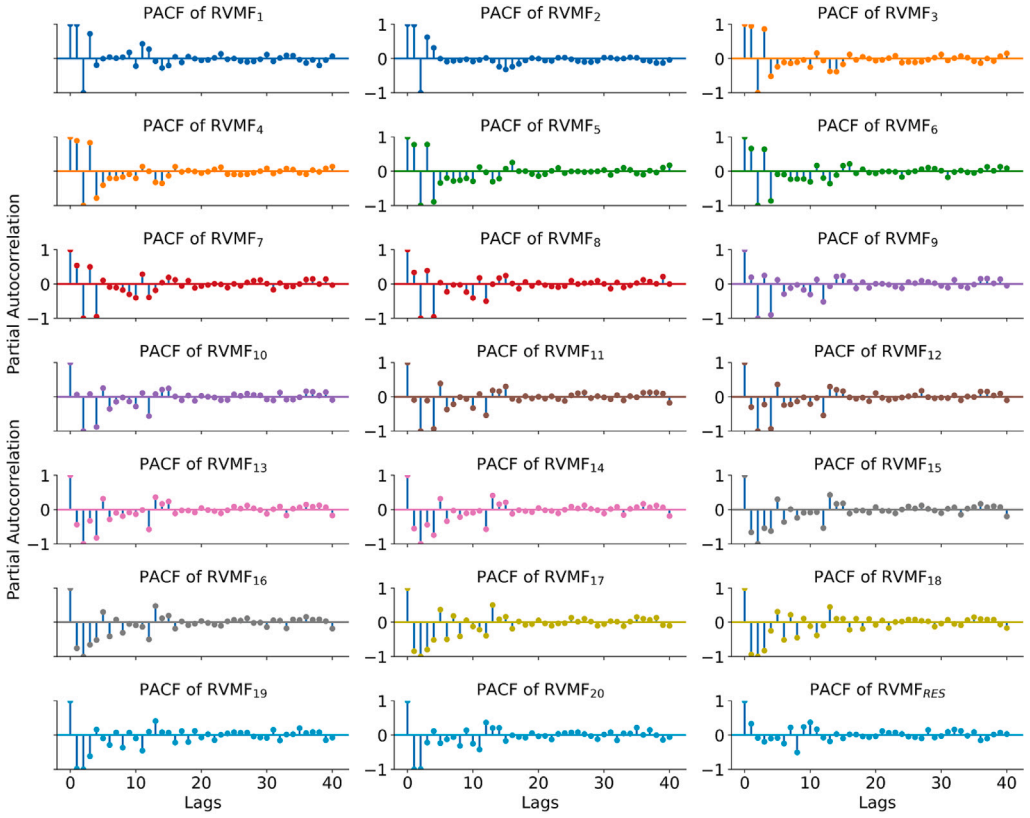


Fig. 6b. PACF of the decomposed residual series showing the most significant lags for the SA, where $RVMF$ is the residual variational mode function.

Table 4
Architecture of D3Net model and other comparative models along with hyperparameter search space for Optuna algorithm. Note: *ReLU* is Rectified Linear Unit activation function, tanh is Hyperbolic Tangent activation function, logistic is Sigmoid activation function, Adaptive Moment Estimation (Adam) is an optimizer and gbtree is Gradient Boosting Tree.

Model	Parameters	Search range	Model	Parameters	Search range
TabNet Model	The number of decision steps, essentially the number of features to be used.	trial.suggest _int("n_d", 8, 64)	Convolution Neural Network integrated with LSTM	Filter 1 (CNN)	('Filter1', range(50,120,5))
	The number of attention steps, related to the model's feature selection.	trial.suggest _int("n_a", 8, 64)		Filter 2 (CNN)	('Filter2', range(50,100,5))
	The number of steps in the decision process.	trial.suggest _int("n_steps", 3, 10)		Filter 3 (CNN)	('Filter3', range(20,80,5))
	The relaxation parameter that controls the sparsity of the feature selection.	trial.suggest _float("gamma", 1.0, 2.0)		Filter 4 (CNN)	('Filter4', range(20,50,5))
	Coefficient for the sparsity regularization.	trial.suggest _float("lambda_sparse", 1e−6, 1e−3, log=True)		LSTM cell 1	('Units 1', range(50,100,5))
	Number of independent Gated Linear Units (GLU) layers.	trial.suggest _int("n_ independent", 1, 5)		Epochs (CNN)	[1000]
	Number of shared GLU layers.	trial.suggest _int("n_shared", 1, 5)		Activation function	[ReLU]
	Parameters for the optimizer, typically Adam optimizer parameters.	lr = trial.suggest _float('lr', 1e−5, 1e−1, log=True)		Solver	['Adam']
	Optimizer	Adam		Batch Size	('Batch_Size', range(50,1500,200))
Deep Neural Network (DNN)	Hiddenneuron 1	('Units1', range(50,120,5))	Long Short Term Memory Network (LSTM)	LSTM cell 1	('Units 1', range(50,100,5))
	Hiddenneuron 2	('Units2', range(50,100,5))		LSTM cell 2	('Units 2', range(50,80,5))
	Hiddenneuron 3	('Units3', range(50,80,5))		Activation function	[ReLU]
	Batch Size	('Batch_Size', range(50,1500,50))		Epochs	[1000]
	Solver	['Adam']		Drop rate	('drop_rate', range(0,0.5,0.1))
	Epochs	[1000]		Batch Size	('Batch_Size', range(50,1500,200))
Random Forest Regression (RFR)	The maximum depth of the tree.	('max_depth', range(1,20,1))	eXtreme Gradient Boosting (XGB)	Booster Type	gbtree
	The number of t rees in the forest.	('n_estimators', range(5,100,2))		Step size shrinkage used in update to prevent overfitting.	('eta', range(0.1,0.9,0.1))
	Minimum number of samples to split an internal node	('min_samples _split', range(2,100,1))		The maximum depth of the tree.	('max_depth', range(1,20,1))
	The number of features to consider when looking for the best split.	['auto', 'sqrt', 'log2']		The number of trees in the forest.	('n_estimators', range(5,100,2))
Multi Layer Perceptron (MLP)	Hidden neuron	[50,60,70, 80,90,100]			
	Activation function	['ReLU','logistic','tanh']			
	Learning rate	[0.001,0.002, 0.005,0.006]			
	Solver	['Adam']			

Determination (R^2), which measures the proportion of variance in the dependent variable that is predictable from the independent variables. The RMSE and Mean Absolute Error (MAE) were used to quantify the differences between predicted and actual values, with RMSE placing greater emphasis on larger errors than MAE. The Legates and McCabe's Efficiency Index (LM) assessed the predictive power and accuracy of the model. Additionally, the Nash–Sutcliffe Efficiency (NS) was employed to evaluate the general predictive skill of the models, while Willmott's Index of Agreement (WI) measured the degree to which predictions were error-free. For relative error measurements, the Relative Root Mean Square Error ($RRMSE$) and Relative Mean Absolute Error ($RMAE$) scaled the RMSE and MAE by the mean of observed values, respectively. The Kling–Gupta Efficiency (KGE) combined correlation, bias, and variability to provide a comprehensive measure of model performance. The Average Percentage Bias (APB) indicated the average bias in percentage terms, while the Symmetric Mean Absolute Percentage Error ($sMAPE$) offered a normalized measure of forecast accuracy. Finally, Theil's Inequality Coefficient (TIC) compared the accuracy of forecasts with naïve predictions, providing insight into the relative quality of the prediction model. Together, these metrics ensured a thorough evaluation of model performance, capturing various aspects of prediction accuracy and reliability. Mathematically, these metrics can be represented as:

1. Coefficient of Determination (R^2):

$$R^2 = \left(\frac{\sum_{i=1}^{N} (EP^a - \langle EP^a \rangle)(EP^p - \langle EP^p \rangle)}{\sqrt{\sum_{i=1}^{n} (EP^a - \langle EP^a \rangle)^2} \sqrt{\sum_{i=1}^{n} (EP^p - \langle EP^p \rangle)^2}} \right)^2, \quad (19)$$

2. Wilmott's Index (WI):

$$WI = 1 - \frac{\sum_{i=1}^{N} (EP^a - EP^p)^2}{\sum_{i=1}^{N} (|EP^p - \langle EP^a \rangle| + |(EP^o - \langle EP^p \rangle|)^2}, \quad (20)$$

3. Nash–Sutcliffe Index (NS):

$$NS = 1 - \frac{\sum_{i=1}^{N} (EP^a - EP^p)^2}{\sum_{i=1}^{N} (EP^a - \langle EP^a \rangle)^2}, \quad (21)$$

4. Legates and McCabe Index (LM):

$$LM = 1 - \frac{\sum_{i=1}^{N} |EP^p - EP^a|}{\sum_{i=1}^{N} |EP^a - \langle EP^a \rangle|}, \quad (22)$$

5. Kling–Gupta Efficiency (KGE):

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle EP^p \rangle}{\langle EP^a \rangle} - 1\right)^2 + \left(\frac{CV^p}{CV^a}\right)^2} \quad (23)$$

6. Root Mean Square Error (RMSE):

$$RMSE(AUD/MWh) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (EP^p - EP^a)^2}, \quad (24)$$

7. Mean Absolute Error (MAE):

$$MAE(AUD/MWh) = \frac{1}{N} \sum_{i=1}^{N} |EP^p - EP^a|, \quad (25)$$

8. Relative Root Mean Square Error($RRMSE$):

$$RRMSE(\%) = \frac{RMSE}{EP^a} \times 100\% \quad (26)$$

9. Relative Mean Absolute($RMAE$):

$$RMAE(\%) = \frac{MAE}{EP^a} \times 100\% \quad (27)$$

10. Symmetric Mean Absolute Percentage Error ($sMAPE$):

$$sMAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|EP^a - EP^p|}{(|EP^a| + |EP^p|)/2}, \quad (28)$$

11. Theil's Inequality Coefficient (TIC):

$$TIC = \frac{\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (EP^p - EP^a)^2}}{\left(\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (EP^a)^2} + \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (EP^p)^2}\right)} \quad (29)$$

12. Absolute Percentage Bias (APB):

$$APB(\%) = \left| \frac{\sum_{i=1}^{n} (EP^a - EP^p)}{\sum_{i=1}^{n} EP^a} \right| \cdot 100, \quad (30)$$

Notable EP^a and EP^p represent actual and predicted half-hourly EP while $\langle EP^a \rangle$ and $\langle EP^p \rangle$ represent actual and predicted mean EP, N = number of tested data points, and CV is the Coefficient of Variation.

This study also utilized Persistence Skill Score (PSS) metrics to evaluate their performance against a persistence-based approach. The PSS metric is defined in Eq. (31), where 'M' represents the RMSE of the predictions made by the models under study, and 'P' corresponds to the RMSE of the persistence model. The persistence model assumes that the EP prediction for the next time step is equal to the EP at the current time step.

$$PSS = 1 - \frac{RMSE_M}{RMSE_P} \quad (31)$$

Additionally, this study employs the Global Performance Indicator (GPI) technique, as described by Ref. [84], to consolidate the values of all twelve indicators (Eq. (19), (20) and (31)) into a single comprehensive measure. The GPI formula incorporates a constant α, where $\alpha = -1$ for metrics to be maximized (R^2, LM, NS, WI, KGE), and $\alpha = 1$ for metrics to be minimized ($RRMSE$, $RMAE$, RMSE, MAE, APB, $sMAPE$, TIC). Each statistical indicator's normalized value is denoted as n_{ik}, with m_k representing the median of the normalized statistical indicator j across all models (where k ranges from 1 to 12). Higher GPI values indicate superior model performance.

$$GPI = \sum_{k=1}^{12} \alpha_k \left(m_k - n_{ik}\right), \quad (-\infty < GPI < +\infty) \quad (32)$$

Furthermore, the comparison of model performance was conducted using three statistical tests: the Diebold-Marino test (DW) [85], Giacomini–White (GW) test [86], and Wilcoxon signed-rank test (WSR) [87]. The DW test evaluates whether one prediction model significantly outperforms another across different prediction horizons by comparing their mean squared prediction errors (MSE) using a $t-test$ adjusted for autocorrelation. The GW test extends beyond the DM test by incorporating a statistic that evaluates the significance of differences in prediction accuracy, accounting for parameter estimation uncertainty through Conditional Predictive Ability (CPA). Meanwhile, the WSR is a non-parametric method that examines the distributions of paired differences between observed and predicted values. It ranks absolute differences, computes sums of ranks for positive and negative differences, and determines if there exists a systematic difference between observed and predicted values. These tests offer robust frameworks for assessing the statistical significance of model performance, aiding analysts and researchers in making informed decisions regarding the suitability and reliability of prediction models across diverse applications.

3.8. Explainability of model outcomes

Finally, advanced model-agnostic tools based on eXplainable Artificial Intelligence (xAI) were leveraged to elucidate the predictions generated by the newly developed D3Net model across two different states. To uncover local explainability, the LIME algorithm was utilized. LIME assesses the effect of each predictor variable in the EP simulation system on an instance basis, offering insights into specific predictions. For broader, global interpretability, the SHAP method was employed. SHAP provides explanations for individual predictions and quantifies the impact of each predictor variable throughout the entire dataset, giving a thorough understanding of the model's behavior and decision-making. These approaches collectively enhance transparency in model outputs, aiding stakeholders in understanding and trusting the D3Net model's performance in practical applications.

4. Results and discussion

The performance of the D3Net model and its individual counterparts was evaluated through a comparative analysis using R^2, RMSE, and MAE metrics. Table 5 provides a numerical summary of the error statistics for all models across several performance measures. During the testing phase, the D3Net model demonstrated superior performance, achieving the lowest error values with an RMSE of 4.629 AUD/MWh and an MAE of 3.051 AUD/MWh for SA, and an RMSE of 4.133 AUD/MWh and an MAE of 2.562 AUD/MWh for TAS. In contrast, the standalone CNN model yielded the poorest results, with an RMSE of $\approx$ 29.46 AUD/MWh and an MAE of $\approx$ 14.26 AUD/MWh for SA, and an RMSE of $\approx$ 29.45 AUD/MWh and an MAE of $\approx$ 13.158 AUD/MWh for TAS during the testing phase. Additionally, the D3Net model achieved the highest R^2 values during the testing period ($R^2 \approx 0.991$ for SA and $R^2 \approx 0.985$ for TAS), further validating its effectiveness for EP prediction. Models that used STL decomposition also outperformed other standalone models during the testing phases. These findings suggest that decomposition-based models are more adept at predicting the

Table 5
Testing phase performance of the D3Net model against the counterpart models in terms of Coefficient of Determination (R^2), Root Mean Square Error ($RMSE$; AUD/MWh) and Mean Absolute Error (MAE; AUD/MWh) in predicting half-hourly EP for South Australia and Tasmania.

Model	South Australia			Tasmania		
	R^2	$RMSE$	MAE	R^2	$RMSE$	MAE
D3Net	**0.991**	**4.629**	**3.051**	**0.985**	**4.133**	**2.562**
STL-MLP	0.719	25.197	13.043	0.312	27.909	13.263
STL-RF	0.714	25.43	12.646	0.297	28.208	13.42
STL-CNN	0.712	25.501	13.059	0.278	28.6	13.567
STL-TabNet	0.711	25.555	13.454	0.307	28.015	13.88
STL-CLSTM	0.705	25.802	14.08	0.285	28.454	13.339
STL-LSTM	0.701	25.997	13.585	0.295	28.264	13.434
TabNet	0.693	26.338	13.346	0.28	28.566	12.846
STL-DNN	0.684	26.722	14.232	0.299	28.168	13.439
MLP	0.682	26.799	13.189	0.282	28.526	12.524
RFR	0.668	27.386	13.627	0.241	29.327	12.622
LSTM	0.659	27.748	13.33	0.233	29.475	13.53
CLSTM	0.653	27.972	13.782	0.229	29.557	13.625
DNN	0.618	29.355	14.164	0.234	29.45	13.359
CNN	0.616	29.46	14.26	0.234	29.45	13.158

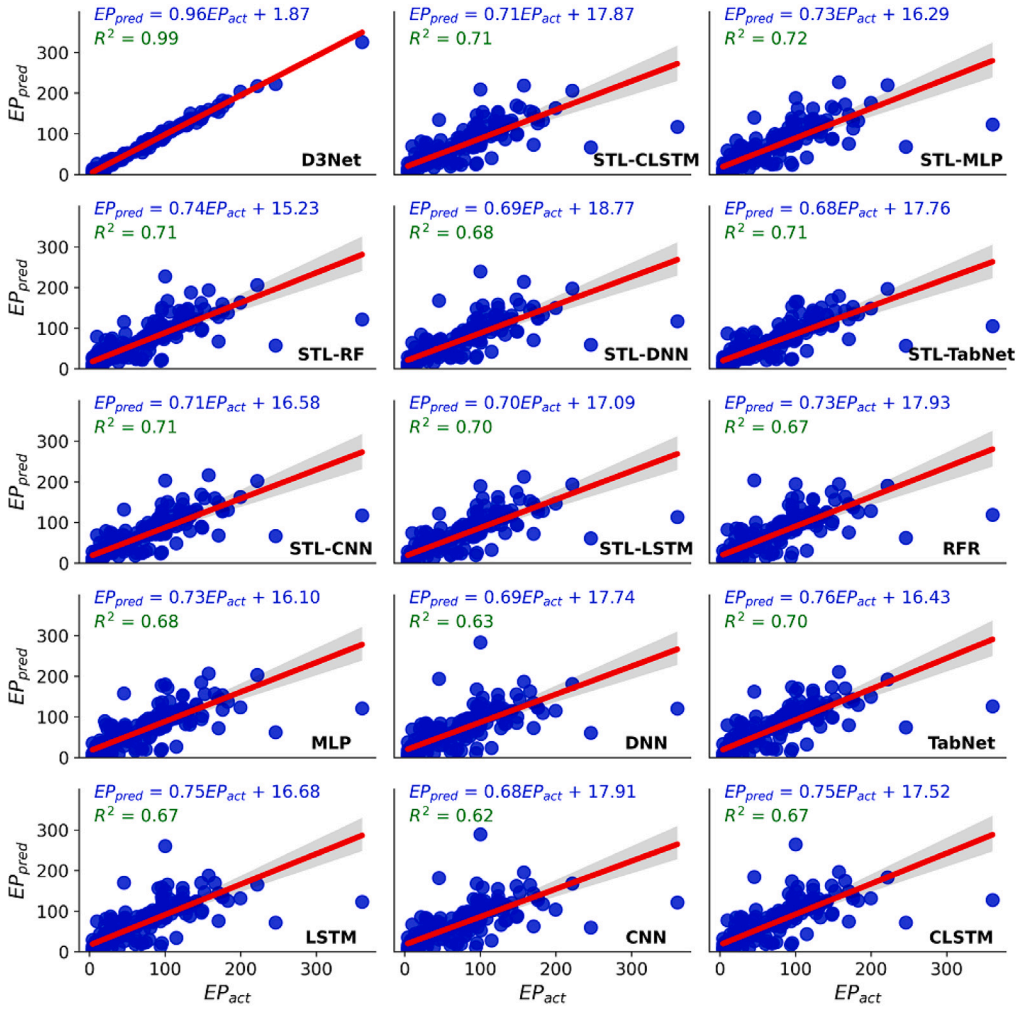


Fig. 7a. Scatter plot of predicted vs. observed half-hourly EP of a SA using the proposed D3Net model and comparing models. A least square regression line and coefficient of determination (R^2) with a linear fit equation are shown in each sub-panel.

non-linear and dynamic characteristics of EP, offering superior non-linear mapping capabilities. The exceptional performance of the D3Net model can be attributed to the use of STL and VMD approaches, which effectively capture the intricate linear, non-stationary, and non-linear features present in EP data. By decomposing complex data into more manageable, stationary sub-components, these techniques facilitate easier handling by the prediction model. Consequently, integrating MLP, RF, and TabNet with signal decomposition techniques significantly enhanced the overall prediction ability of the D3Net model. The D3Net model successfully extracts crucial information from EPdata and accurately mimics its complex and stochastic behavior to predict future values. Overall, the proposed D3Net model delivered highly satisfactory results in predicting half-hourly EP, underscoring the applicability of the developed models for EP prediction.

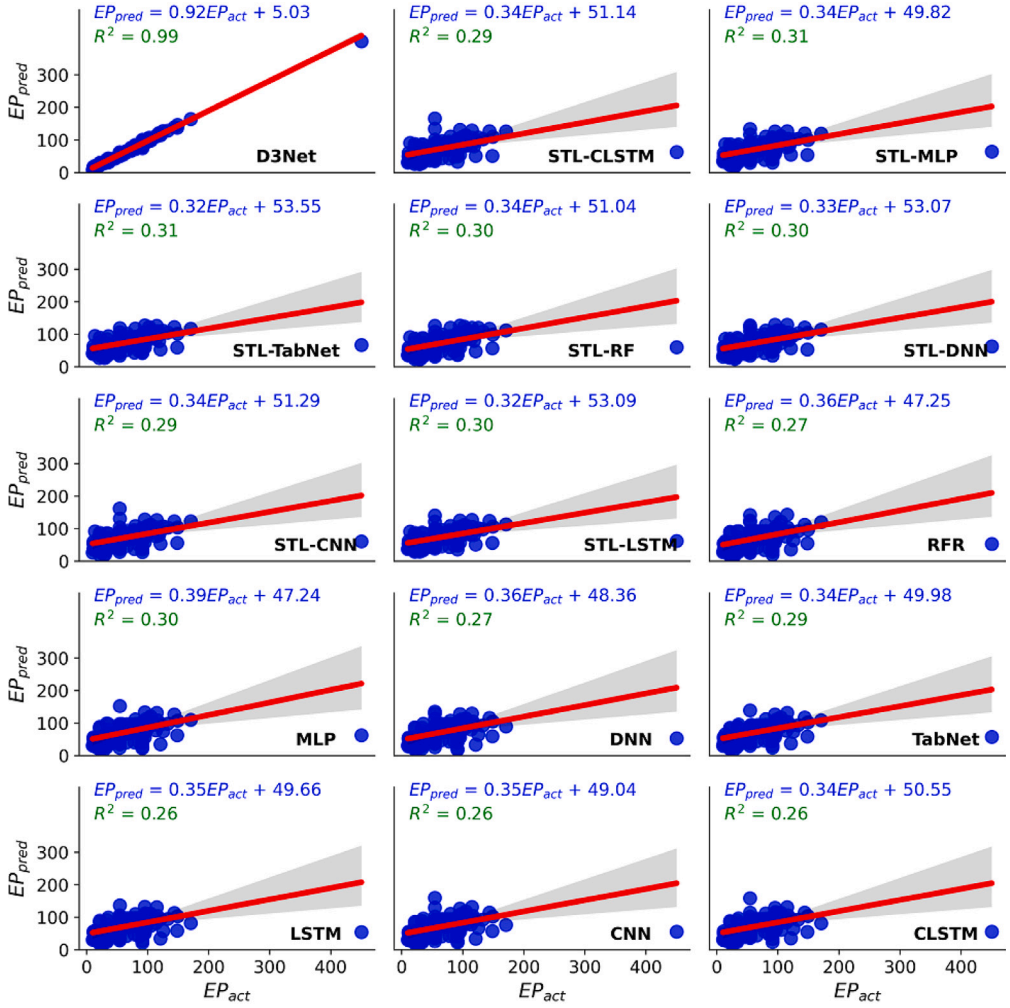


Fig. 7b. Same as Fig. 7a for TAS.

Table 6
Accuracy measures of half-hourly EP prediction for D3Net and comparative models in terms of Willmott's Index (WI), Nash–Sutcliffe Efficiency (NS), Legates–McCabe's Index (LM).

Model	South Australia			Tasmania		
	WI	*NS*	*LM*	*WI*	*NS*	*LM*
D3Net	**0.998**	**0.991**	**0.961**	**0.996**	**0.985**	**0.946**
STL-MLP	0.913	0.719	0.823	0.692	0.312	0.68
STL-RF	0.913	0.714	0.829	0.69	0.297	0.68
STL-CNN	0.909	0.712	0.821	0.683	0.278	0.676
STL-TabNet	0.906	0.711	0.813	0.679	0.307	0.66
STL-CLSTM	0.907	0.705	0.806	0.69	0.285	0.683
STL-LSTM	0.905	0.701	0.813	0.678	0.295	0.673
TabNet	0.91	0.693	0.827	0.688	0.28	0.697
STL-DNN	0.899	0.684	0.803	0.682	0.299	0.673
MLP	0.903	0.682	0.825	0.712	0.282	0.72
RFR	0.9	0.668	0.82	0.695	0.241	0.716
LSTM	0.9	0.659	0.827	0.688	0.233	0.695
CLSTM	0.899	0.653	0.821	0.682	0.229	0.688
DNN	0.881	0.618	0.809	0.69	0.234	0.698
CNN	0.88	0.616	0.807	0.683	0.234	0.698

To gain a more comprehensive understanding of the prediction performance, scatter plots were generated and are illustrated in Figs. 7a and 7b. These scatter plots include the coefficient of determination (R^2) and the equation of the regression line ($y = mx + c$, where, $}m'$ is defined as the gradient, and $}C'$ is denoted as the y-intercept) to further evaluate the developed models. The R^2 value indicates how well the data points fit the regression line, with values closer to 1 signifying better model performance. During the testing phase, the D3Net model exhibited the best regression results, with its scatter plot showing data points closely aligned along the regression line, indicating a near-perfect fit (1:1) for both SA and TAS. This suggests that the D3Net model is highly accurate in predicting the target variable. Following closely behind in performance were the STL decomposed models, which also showed strong regression results, albeit slightly less precise than the D3Net model. In contrast, standalone models, including MLP, DNN, TabNet, LSTM, CNN, and CLSTM, demonstrated poorer performance.

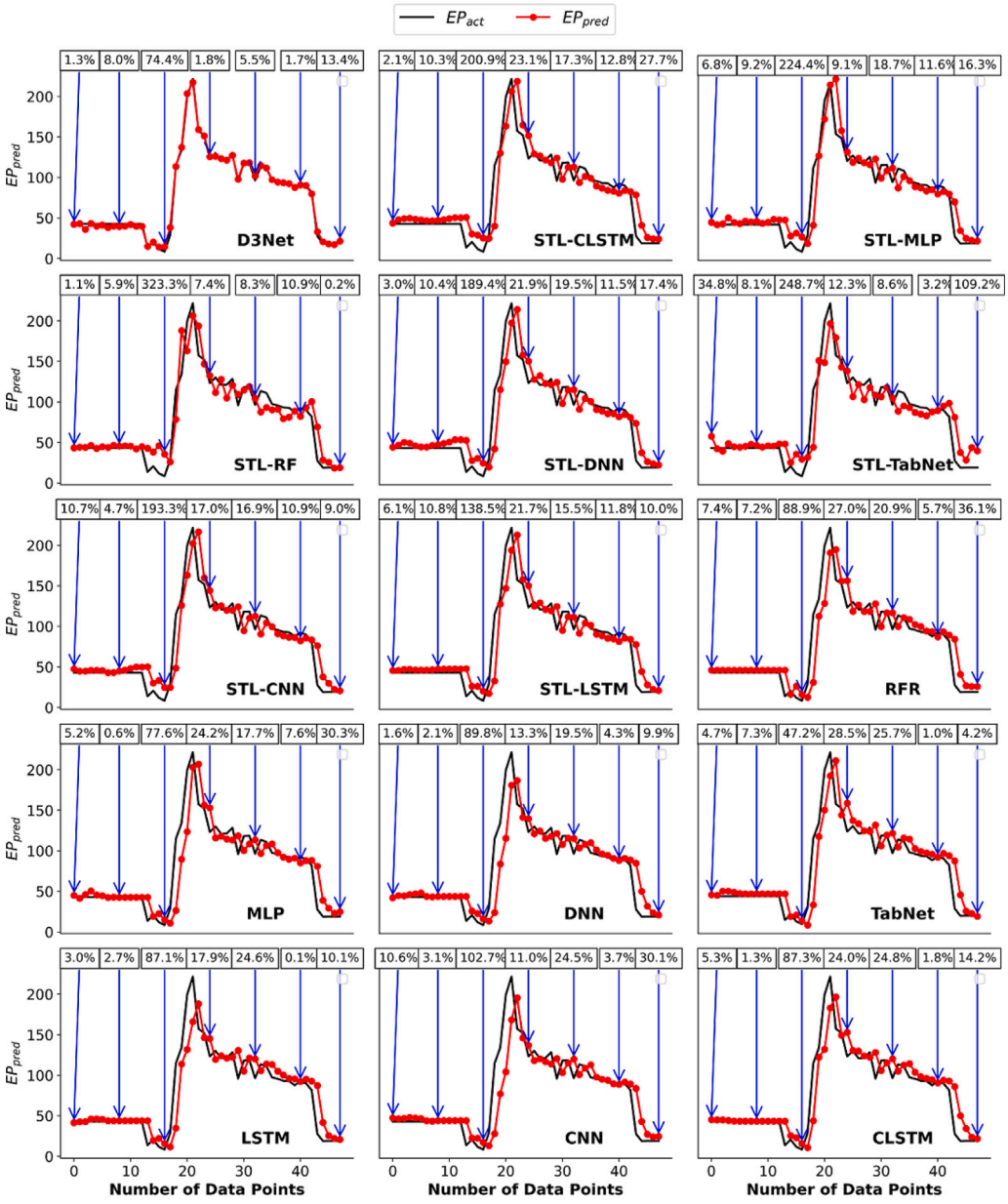


Fig. 8a. Variation trend of the developed models in the testing phase for SA for predicting half-hourly EP. The relative error are also shown in the plot. Note:- For the purpose of brevity only one day trend is shown.

Their scatter plots displayed highly dispersed data points around the regression line, indicating less accurate predictions. This dispersion signifies that these models are less effective at capturing the underlying patterns in the data, leading to larger prediction errors. Among these, the RFR model showed the poorest goodness-of-fit results, with its scatter plot displaying the highest level of data point dispersion, indicating the lowest prediction accuracy. The superior performance of the D3Net model can be attributed to its advanced decomposition prediction algorithm, which incorporates VMD along with STL. This algorithm enhances the model's learning ability by breaking down the complex time series data into simpler, more manageable components—seasonal, trend, and residual. By decomposing the seasonal and residual components of STL further using VMD, the D3Net model effectively captures the intricate patterns and non-linear relationships present in the data.

Comparisons of the actual and predicted EP testing data in terms of WI, NS, and LM are further examined in Table 6. These metrics provide a comprehensive evaluation of the prediction models' accuracy and reliability. For both SA and TAS, the D3Net model outperformed all other models, achieving the highest values in all three performance metrics. Specifically, the D3Net model recorded WI values of ≈ 0.998 for SA and ≈ 0.996 for TAS, indicating near-perfect agreement between the actual and predicted data. Similarly, the NS values for the D3Net model were ≈ 0.991 for SA and ≈ 0.985 for TAS, demonstrating its exceptional predictive accuracy. The LM values were equally impressive, with the D3Net model scoring ≈ 0.961 for SA and ≈ 0.946 for TAS, further confirming its superior performance. In contrast, the standalone CNN model yielded the lowest performance metrics for SA, with the lowest $WI \approx 0.88$, $NS \approx 0.616$, and $LM \approx 0.807$ values among the evaluated models. This indicates that the CNN model struggled to accurately predict the EP data for this region. For TAS, the STL-LSTM model produced the lowest $WI \approx 0.679$, and $LM \approx 0.66$ values, suggesting it was the least effective in capturing the patterns in the EP data for this region. These results highlight the robustness and reliability of the D3Net model, especially when compared to both standalone and other decomposition based models. The high values

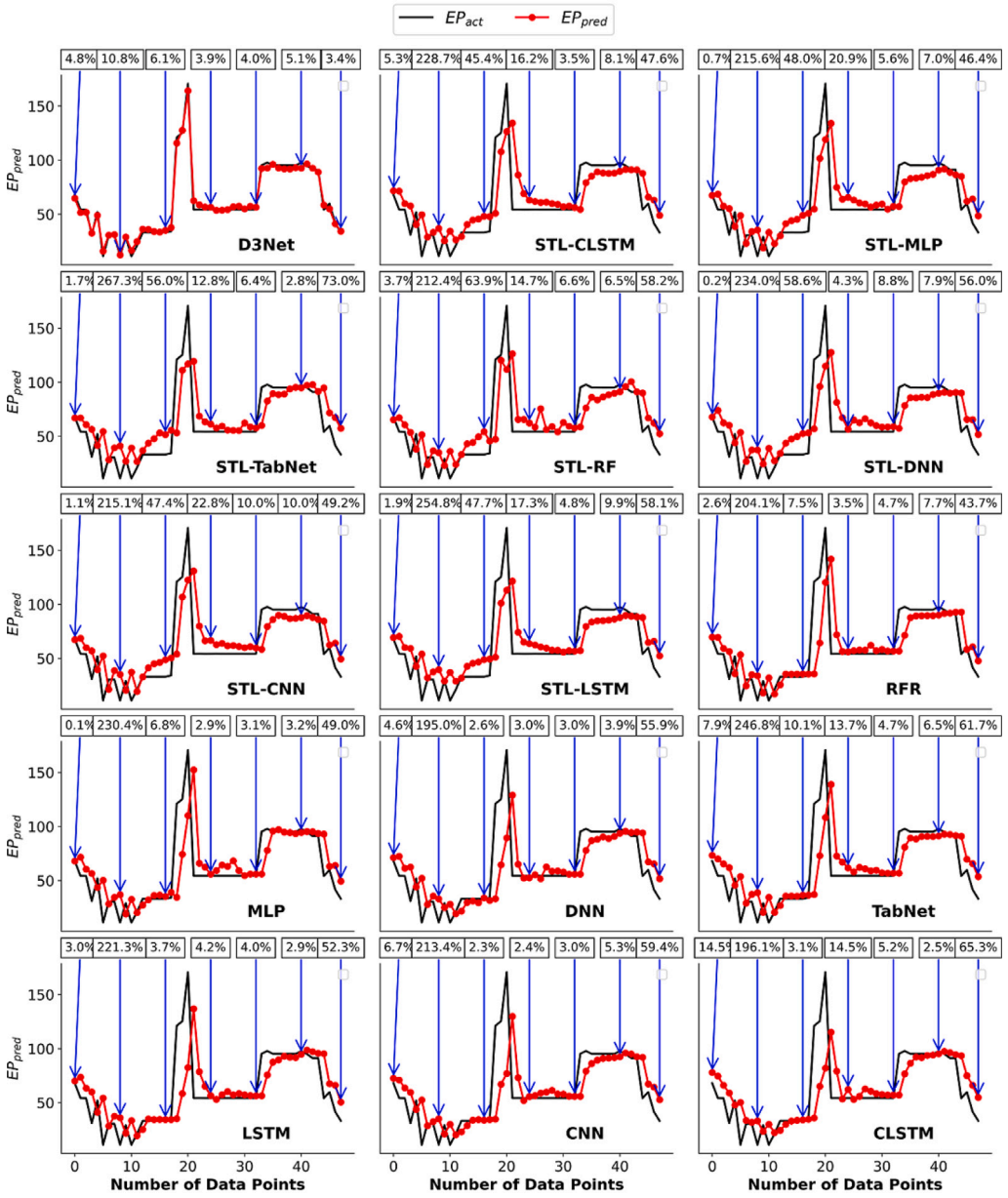


Fig. 8b. Variation trend of the developed models in the testing phase for TAS for predicting half-hourly EP. The relative error are also shown in the plot. Note:- For the purpose of brevity only one day trend is shown.

of WI, NS, and LM achieved by the D3Net model underscore its capability to closely mimic the actual EP data, making it a highly effective tool for EP prediction. The poor performance of the CNN and STL-LSTM models in SA and TAS, respectively, further emphasizes the importance of model selection and highlights the benefit of using advanced decomposition techniques, such as those employed in the D3Net model, to further improve predictive accuracy.

Figs. 8a and 8b illustrate the trend and relative error of all models over a one-day period (48 half-hourly data points) for SA and TAS. While all models aimed to replicate the actual EP data pattern, the D3Net model notably excelled by closely matching the actual data with low relative errors. In contrast, the other models showed greater variations and higher relative errors, deviating more from the actual data. These trends demonstrate that the D3Net model's ability to manage non-linear and non-stationary data is superior to that of the other models. The results confirm that the D3Net model, utilizing STL and VMD techniques, excels inEP prediction accuracy, underscoring the effectiveness of signal decomposition techniques.

Further evaluation of the D3Net model is conducted using various metrics, including Relative RMSE ($RRMSE$), Relative MAE ($RMAE$), Kling–Gupta Efficiency (KGE), Absolute Percentage Bias (APB), Symmetric Mean Absolute Percentage Error ($sMAPE$), and Theil's Inequality Coefficient (TIC), as detailed in Table 7 and 8. Across these metrics, the D3Net model consistently demonstrates superior performance. Specifically, the D3Net model achieves the lowest $RRMSE$ values ($\approx 7.90\%$ for SA, $\approx 5.40\%$ for TAS) and the lowest RMAE values ($\approx$ 5.20% for SA, $\approx$ 3.40% for TAS). It also records the lowest APB values ($\approx$ 11.509% for SA, $\approx$ 4.272% for TAS), the lowest $sMAPE$ values ($\approx$ 0.1% for SA, $\approx$ 0.041% for TAS), and the lowest TIC values ($\approx$ 0.031% for SA, $\approx$ 0.025% for TAS). In addition to these, the D3Net model attains the highest KGE values ($\approx$ 0.968% for SA, $\approx$ 0.927% for TAS), highlighting its robust predictive accuracy and reliability. These findings align with earlier results, reinforcing the D3Net model's exceptional ability to predict EP data accurately. The model's effectiveness is primarily attributed to the incorporation of Seasonal-Trend decomposition using LOESS (STL) and Variational Mode Decomposition

Table 7
Relative Error (Relative RMSE and Relative MAE) evaluation results of D3Net vs. comparative models in predicting half-hourly EP for South Australia and Tasmania.

Model	South Australia		Tasmania	
	RRMSE	*RMAE*	*RRMSE*	*RMAE*
D3Net	**7.90%**	**5.20%**	**5.40%**	**3.40%**
STL-MLP	43.00%	22.20%	36.50%	17.40%
STL-RF	43.30%	21.60%	36.90%	17.60%
STL-CNN	43.50%	22.30%	37.40%	17.80%
STL-TabNet	43.60%	22.90%	36.70%	18.20%
STL-CLSTM	44.00%	24.00%	37.20%	17.50%
STL-LSTM	44.30%	23.20%	37.00%	17.60%
TabNet	44.90%	22.80%	37.40%	16.80%
STL-DNN	45.60%	24.30%	36.90%	17.60%
MLP	45.70%	22.50%	37.30%	16.40%
RFR	46.70%	23.20%	38.40%	16.50%
LSTM	47.30%	22.70%	38.60%	17.70%
CLSTM	47.70%	23.50%	38.70%	17.80%
DNN	50.00%	24.10%	38.50%	17.50%
CNN	50.20%	24.30%	38.50%	17.20%

Table 8
Comparison of prediction results of the proposed model (D3Net) and other comparative models across various metrics for South Australia and Tasmania. Metrics include Kling–Gupta Efficiency (KGE), Absolute Percentage Bias (APB), Symmetric Mean Absolute Percentage Error ($sMAPE$), and Theil Inequality Coefficient (TIC).

Model	South Australia				Tasmania			
	APB	*KGE*	*sMAPE*	*TIC*	*APB*	*KGE*	*sMAPE*	*TIC*
D3Net	**11.509**	**0.968**	**0.1**	**0.031**	**4.272**	**0.927**	**0.041**	**0.025**
STL-MLP	38.166	0.795	0.282	0.171	25.492	0.41	0.195	0.172
STL-RF	36.531	0.799	0.256	0.173	26.364	0.407	0.198	0.173
STL-CNN	36.76	0.779	0.273	0.174	26.396	0.403	0.199	0.175
STL-TabNet	43.043	0.751	0.297	0.177	28.458	0.388	0.206	0.171
STL-CLSTM	43.086	0.774	0.314	0.176	26.115	0.413	0.194	0.174
STL-LSTM	38.335	0.768	0.29	0.178	27.019	0.385	0.197	0.173
TabNet	35.133	0.809	0.259	0.175	24.588	0.409	0.189	0.176
STL-DNN	44.21	0.762	0.315	0.182	27.01	0.393	0.199	0.172
MLP	34.398	0.788	0.264	0.182	23.452	0.461	0.183	0.174
RFR	38.259	0.786	0.279	0.183	23.293	0.431	0.184	0.181
LSTM	32.968	0.797	0.248	0.185	24.815	0.421	0.194	0.18
CLSTM	38.949	0.794	0.275	0.185	24.846	0.411	0.194	0.181
DNN	36.593	0.754	0.273	0.2	24.015	0.425	0.193	0.181
CNN	37.764	0.751	0.284	0.201	24.006	0.414	0.19	0.181

(VMD) techniques. The use of advanced signal decomposition methods considerably boosts the model's effectiveness in handling intricate, non-linear, and non-stationary data patterns Overall, the comprehensive evaluation metrics underscore the D3Net model's dominance in EP prediction. The results validate the advantages of using sophisticated signal decomposition techniques, establishing the D3Net model as a highly reliable and accurate tool for energy price prediction. Appendix B.1 additionally shows a frequency analysis of the prediction errors.

Fig. 9 provides a comprehensive comparison of all models based on the Persistence Skill Score (PSS) metrics obtained during the testing phases. The D3Net model consistently yielded outstanding results, with high PSS scores of $\approx$ 0.84 and $\approx$ 0.89. In contrast, the DNN and CNN models exhibited the poorest performance. These findings are consistent with the data presented in Tables Table 7, and Table 8, which indicate that the D3Net model predominantly generates prediction errors within the lowest magnitude band, outperforming the persistence model. Consequently, D3Net demonstrates higher accuracy in predicting electricity prices (EP) for the Australian states compared to the other models examined.

Table 9 presents the ranking of the investigated EP prediction models based on the GPI. The model with the greatest GPI value is considered to perform the best. According to the GPI, the D3Net model is the most suitable, ranking first among all developed models with a GPI of $\approx$ 11.068 for SA and $\approx$ 12.20 for TAS. In comparison, the STL-MLP model ranks second, but with significantly lower GPI values of $\approx$ 0.483 for SA and $\approx$ 0.137 for TAS. These results highlight the superior performance of the D3Net model in EP for these regions.

Table 9
Global Performance Indicator(GPI) to compare the various models used in half-hourly EP prediction.

Model	South Australia	Tasmania
	GPI Score	*GPI Score*
D3Net	**11.068**	**12.207**
STL-MLP	0.484	0.138
STL-RF	0.708	−0.035
STL-CNN	0.352	−0.224
STL-Tabnet	−0.262	−0.328
STL-CLSTM	−0.440	−0.031
STL-LSTM	−0.152	−0.177
Tabnet	0.386	0.142
STL-DNN	−0.905	−0.128
MLP	0.094	0.551
RFR	−0.396	0.169
LSTM	−0.042	−0.267
CLSTM	−0.558	−0.371
DNN	−1.385	−0.163
CNN	−1.564	−0.154

To confirm the dominance of the best models, this study conducted the Diebold–Mariano (DM) test on each pair of models. The null hypothesis posits that the two models compared (one from the columns and one from the rows) are equally accurate, while the alternative hypothesis asserts that the column model is more accurate than the row model, based on the RMSE. The results of the DM test, presented in

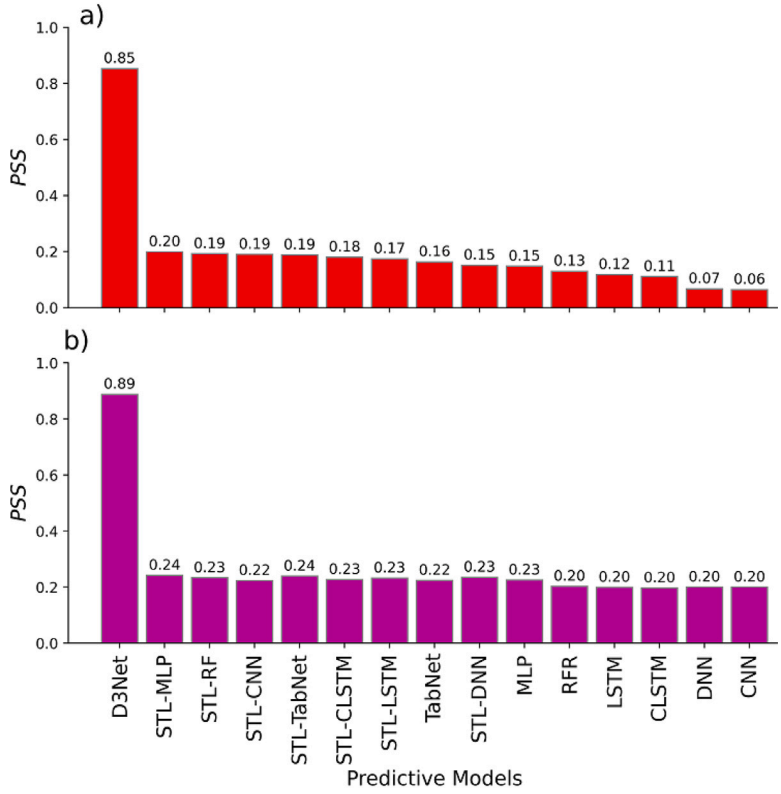


Fig. 9. Bar plot of the model performances in predicting half-hourly EP using the D3Net model against the other models in the testing phase. (a) SA and (b) TAS.

Table 10
Comparison of prediction outcomes using non-parametric Wilcoxon signed-rank test.

Model	South Australia			Tasmania		
	Positive signs ranks sum	Negative signs ranks sum	z-score	Positive signs ranks sum	Negative signs ranks sum	z-score
D3Net	**27912.500**	**30741.500**	**7.628**	**36004.500**	**22307.500**	**7.883**
STL-CLSTM	21353.500	37300.500	5.812	25924.500	32387.500	7.087
STL-MLP	20760.500	37893.500	5.651	28837.500	29474.500	6.129
STL-RF	22725.500	35928.500	6.187	26805.500	31506.500	7.327
STL-DNN	20301.500	38352.500	5.525	25140.500	33171.500	6.873
STL-TabNet	23705.500	34948.500	6.453	24676.500	33635.500	6.745
STL-CNN	21866.500	36787.500	5.952	27656.500	30655.500	7.560
STL-LSTM	22183.500	36470.500	6.039	28137.500	30174.500	7.692
RFR	12545.500	46108.500	3.415	27386.500	30925.500	7.495
MLP	20116.500	38537.500	5.479	22138.500	36173.500	6.058
DNN	18463.500	40190.500	5.029	24719.500	33592.500	6.759
TabNet	13515.500	45138.500	3.680	31651.500	26660.500	7.290
LSTM	13886.500	44767.500	3.784	21366.500	36945.500	5.843
CNN	18915.500	39738.500	5.154	27292.500	31019.500	7.463
CLSTM	13473.500	45180.500	3.670	20804.500	37507.500	5.689

Fig. 11, reveal the $DM-statistic$ for each comparison. A negative DM-statistic suggests that the row model performs statistically better than the column model. The findings for SA and TAS demonstrate that the D3Net model is statistically superior for half-hourly EP prediction at the 5% significance level. Appendix B.2 additionally shows the results of a statistical GW test.

Table 10 compares the outcomes of the developed models with the actual data using the Wilcoxon signed-rank test. All models exhibit z-values greater than 1.96 during testing, leading to the rejection of the null hypothesis (The null hypothesis in the Wilcoxon signed-rank test states that there is no difference between the paired samples.). Notably, the D3Net model demonstrates the best performance, with $z-values$ of ≈ 7.628 for SA and ≈ 7.883 for TAS. These outstanding results for D3Net are consistent with earlier findings and further validate its efficacy in EP prediction.

4.1. Explainability results

SHAP and LIME are efficient tools for interpreting ML and DL models, transforming black-box models into glass-box models. SHAP provides a comprehensive view of feature importance and interactions across the entire dataset, making it ideal for global explanations. In contrast, LIME offers detailed insights into specific predictions, making it perfect for local explanations. By highlighting which features most significantly influence predictions, SHAP and LIME further help identify key drivers and potential areas for model improvement. In this study, the interpretations of each D3Net model (MLP, RFR, and TabNet) were performed using both SHAP and LIME, as detailed below:

1. **Interpretability of MLP model for trend component prediction:** Fig. 11a is a SHAP summary plot or bee swarm, which visually represents the impact of different features on the output

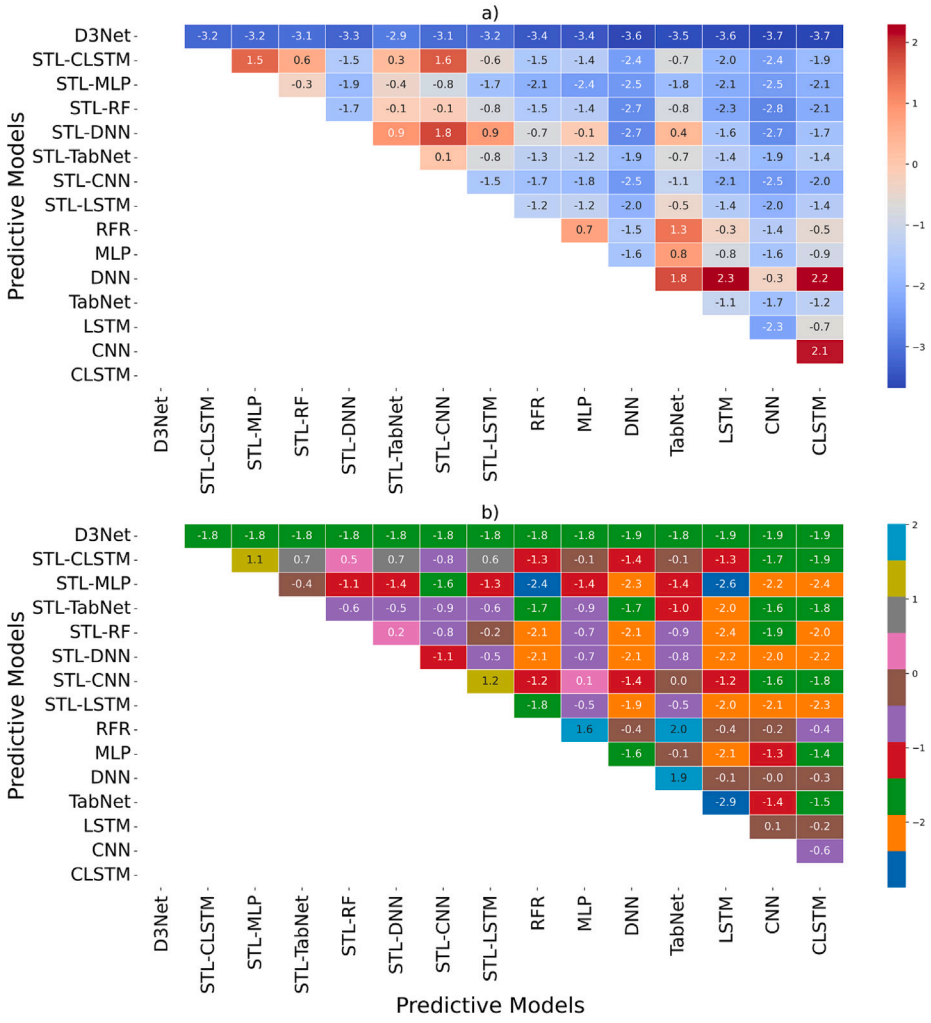


Fig. 10. Comparison of prediction outcomes using statistical DM test. (a) SA and (b) TAS.

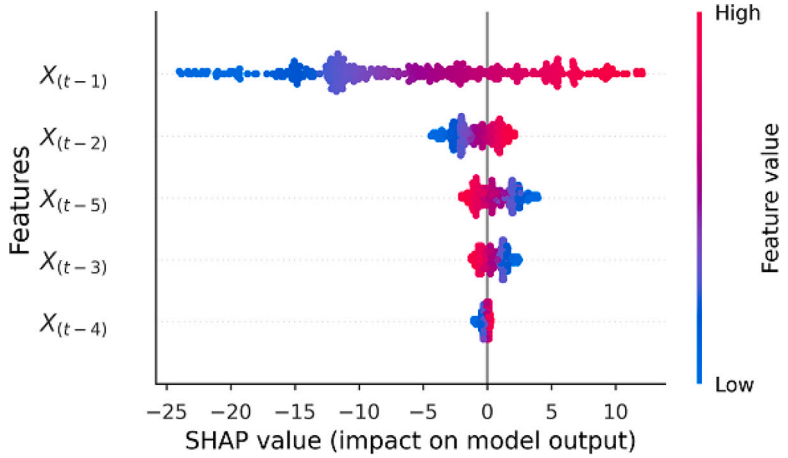


Fig. 11a. SHAP summary bee-swarm plots for the TAS. Note:- This plot is only for Trend prediction using MLP model after STL decomposition. $X_{(t-i)}$ where $i = 0, 1, 2, 3...5$, are the lagged values of trend component. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of a MLP model for trend prediction after STL decomposition off the EP time-series. The plot displays the SHAP values for five lagged features of trend component, denoted as $X_{(t-1)}$, $X_{(t-2)}$, $X_{(t-5)}$, $X_{(t-3)}$, and $X_{(t-4)}$, on the x-axis. The SHAP values quantify the contribution of each feature to the model's prediction for each instance in the dataset. Each dot (blue and red) represents an individual instance's feature value and its corresponding SHAP value, with colors indicating the feature value magnitude (ranging from low values in blue to high values in red). The $x - axis$ indicates the SHAP value, which shows the direction and magnitude of each feature's effect on the model output. For instance, positive SHAP values push the prediction higher, while

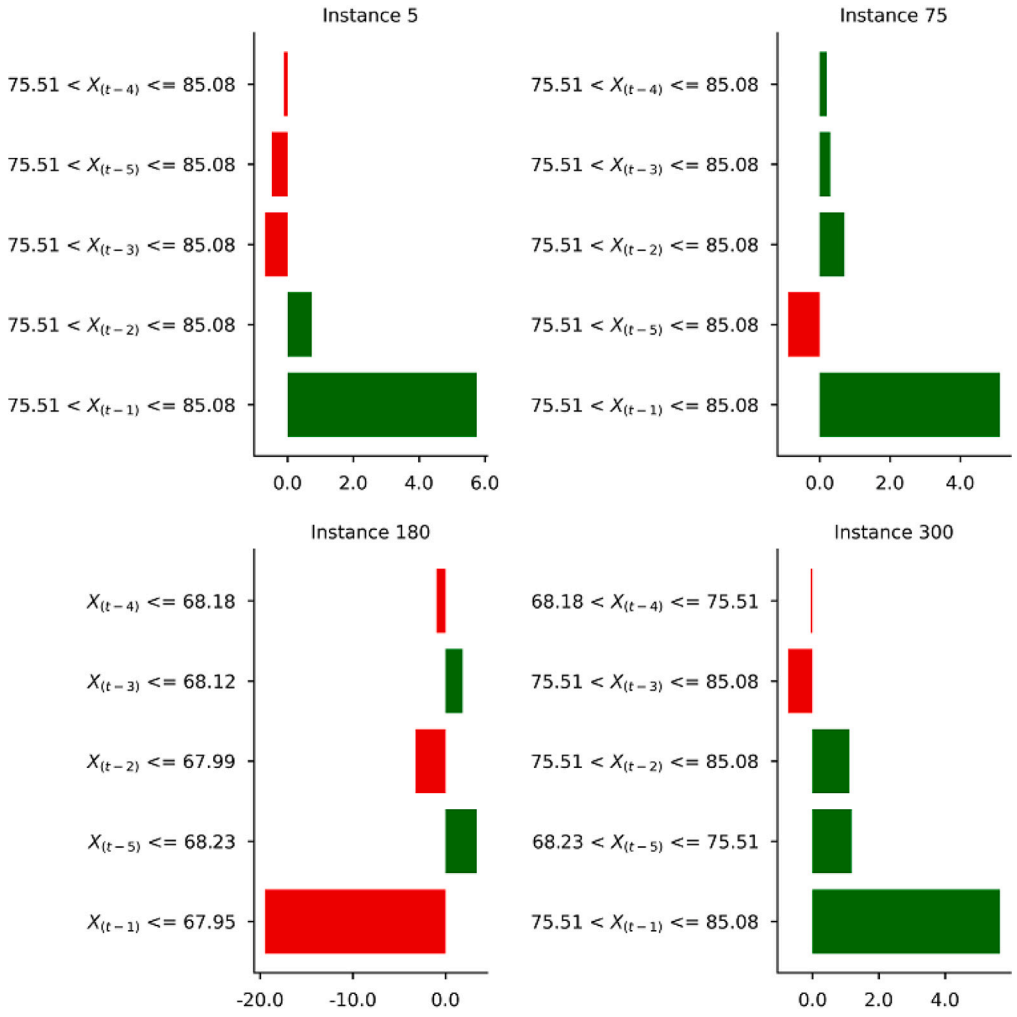


Fig. 11b. LIME explanation bar plots at (i) instance 5, (ii) instance 75, (iii) instance 180, and (iv) instance 300 for the TAS. Where the green bars indicate that the features have a positive impact on the model (increase the model score) and the red bars indicate that the features have a negative impact on the model (decrease the model score). Note:- This plot is only for trend prediction using MLP model after STL decomposition. $X_{(t-i)}$ where $i = 0, 1, 2, 3...5$, are the lagged values of trend component. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

negative values decrease it. The vertical position of the dots along each feature's row shows the spread of SHAP values for that feature, providing insights into the distribution of feature impacts. From the plot, we can infer which features have the most significant influence on the model's predictions. For example, $X_{(t-1)}$ has a wide spread of SHAP values, indicating that it strongly influences the model's output in both positive and negative directions. In contrast, $X_{(t-4)}$ has SHAP values mostly clustered around zero, suggesting a lesser impact on the model's predictions.

Fig. 11b displays a set of LIME decision plots for individual instances in a MLP model for trend component prediction. Each subplot corresponds to a specific instance (Instance 5, Instance 75, Instance 180, and Instance 300) and shows how different features contribute to the model's prediction for that instance. For each subplot, the vertical axis lists the features and their corresponding decision thresholds. The horizontal bars represent the contribution of each feature to the final prediction, with the length of the bar indicating the magnitude of the contribution. Positive contributions are shown in green, while negative contributions are in red. The values at the end of each bar indicate the range of the feature values.

- **Instance 5 and Instance 75**: Both show positive contributions from multiple features, with $X_{(t-1)}$ having the most significant positive impact on the model's prediction.
- **Instance 180**: This instance has a large negative contribution from $X_{(t-1)}$, leading to a significantly lower prediction, with several other features also contributing negatively.
- **Instance 300**: Similar to Instance 5 and Instance 75, this instance has several features contributing positively, with $X_{(t-1)}$ being the most influential positive feature.

2. **Interpretability of RFR model for seasonal component prediction:** As previously mentioned, the seasonal component of the STL decomposition is further decomposed using VMD into 20 VMFs, denoted as SVMF. Each sub-series (SVMF) is predicted using the RFR model. The input for the RFR model consists of the lagged series, denoted as $X_{(t-n)}$, for each SVMF component (where t is the current time step and n is the lag.). Fig. 12a displays the SHAP value bar charts for multiple features across different SVMF sub-series. The SVMF_{Res} represents the residual component after the VMD decomposition of the seasonal component. Each subplot in Fig. 12a highlights the importance of various features for a RFR model. The SHAP values displayed

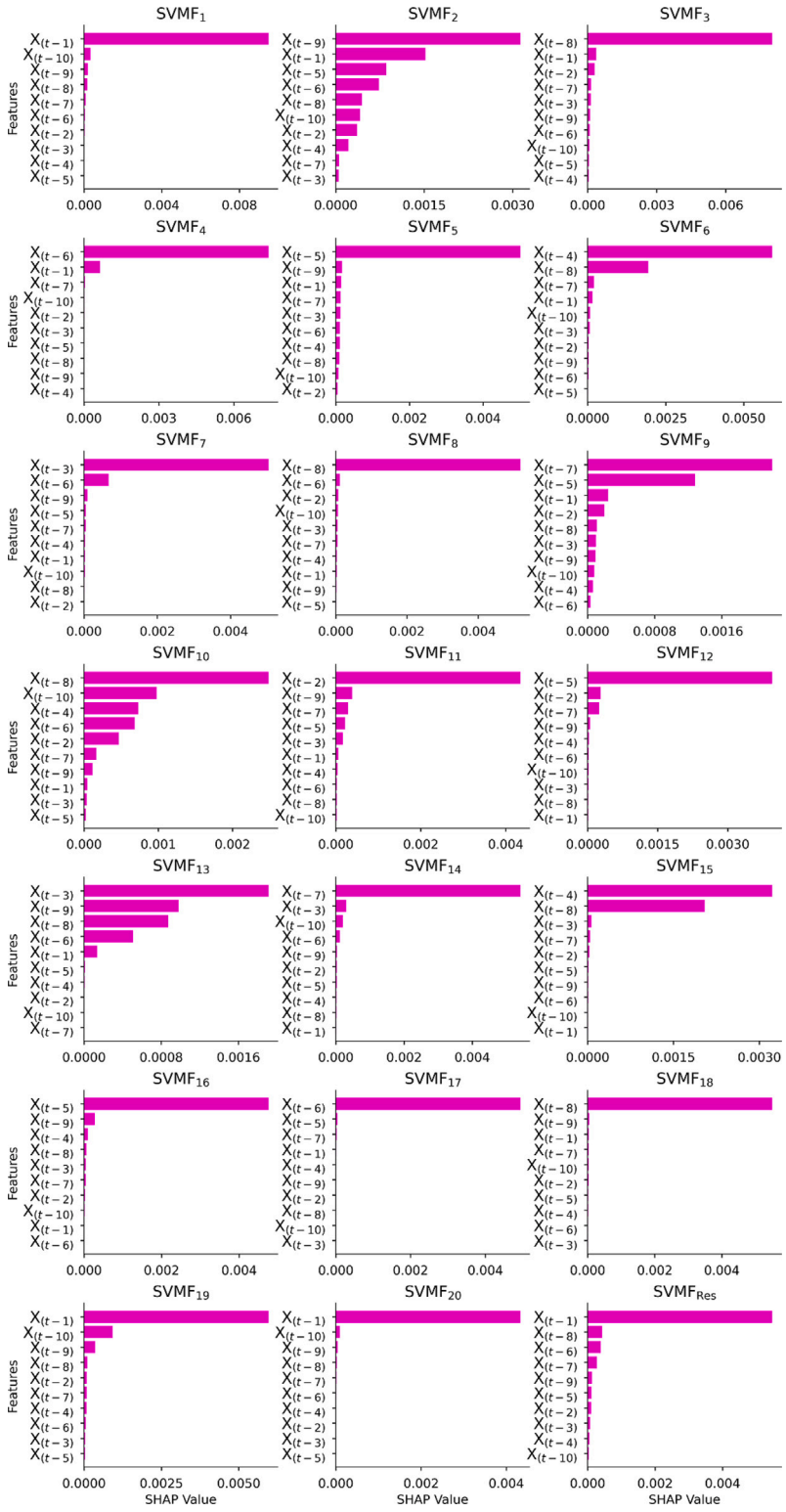


Fig. 12a. SHAP summary bar plots for the TAS. Note:- This plot is only for seasonal prediction using RF model after VMD decomposition. $SVMF$ is the seasonal variation mode function. $X_{(t-i)}$ where $i = 0, 1, 2, 3...10$, are the lagged values of $SVMF$ component.

on the x-axis quantify each feature's contribution to the model's prediction, with higher values signifying greater importance.

- **SVMF$_1$**: In this subplot, the feature $X_{(t-1)}$ has the highest SHAP value, indicating it is the most influential feature for the model's prediction. This suggests that the value of the

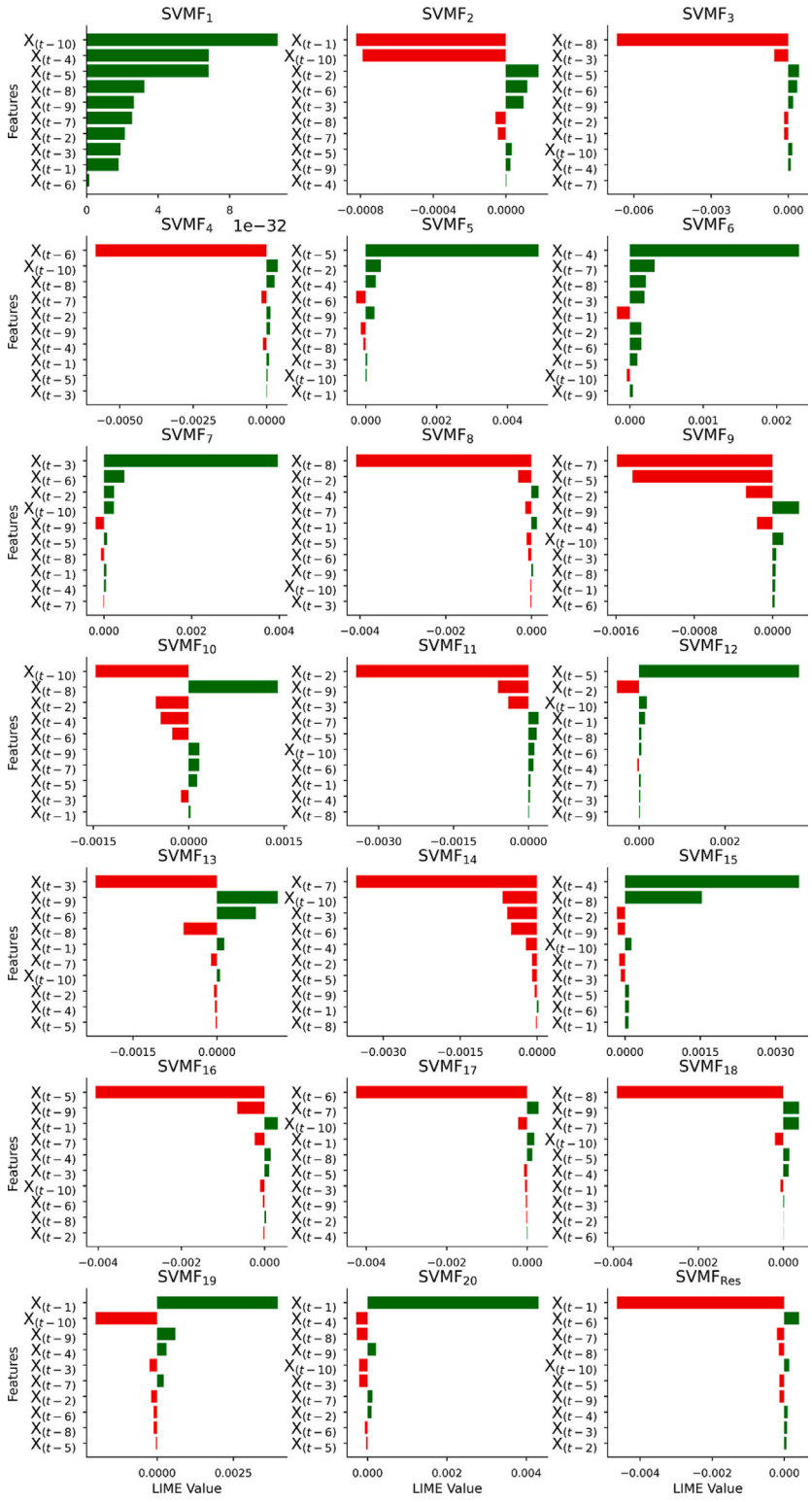


Fig. 12b. LIME explanation bar plots at instance 300 for TAS. Note:- This plot is only for seasonal prediction using RF model after VMD decomposition. $SVMF$ is the seasonal variation mode function. $X_{(t-i)}$ where $i = 0, 1, 2, 3...10$, are the lagged values of $SVMF$ component.

feature at time $t-1$ is crucial for the predictions made by the SVMF_1 model.

- **SVMF_2**: The feature $X_{(t-9)}$ has the highest SHAP value, followed by $X_{(t-1)}$ and $X_{(t-5)}$. This indicates that these features significantly impact the predictions of the SVMF_2 model.
- **SVMF_5**: The feature $X_{(t-5)}$ has a substantial SHAP value, making it the most influential feature for this model. Other features like $X_{(t-9)}$ and $X_{(t-1)}$ also contribute to the prediction but to a lesser extent.

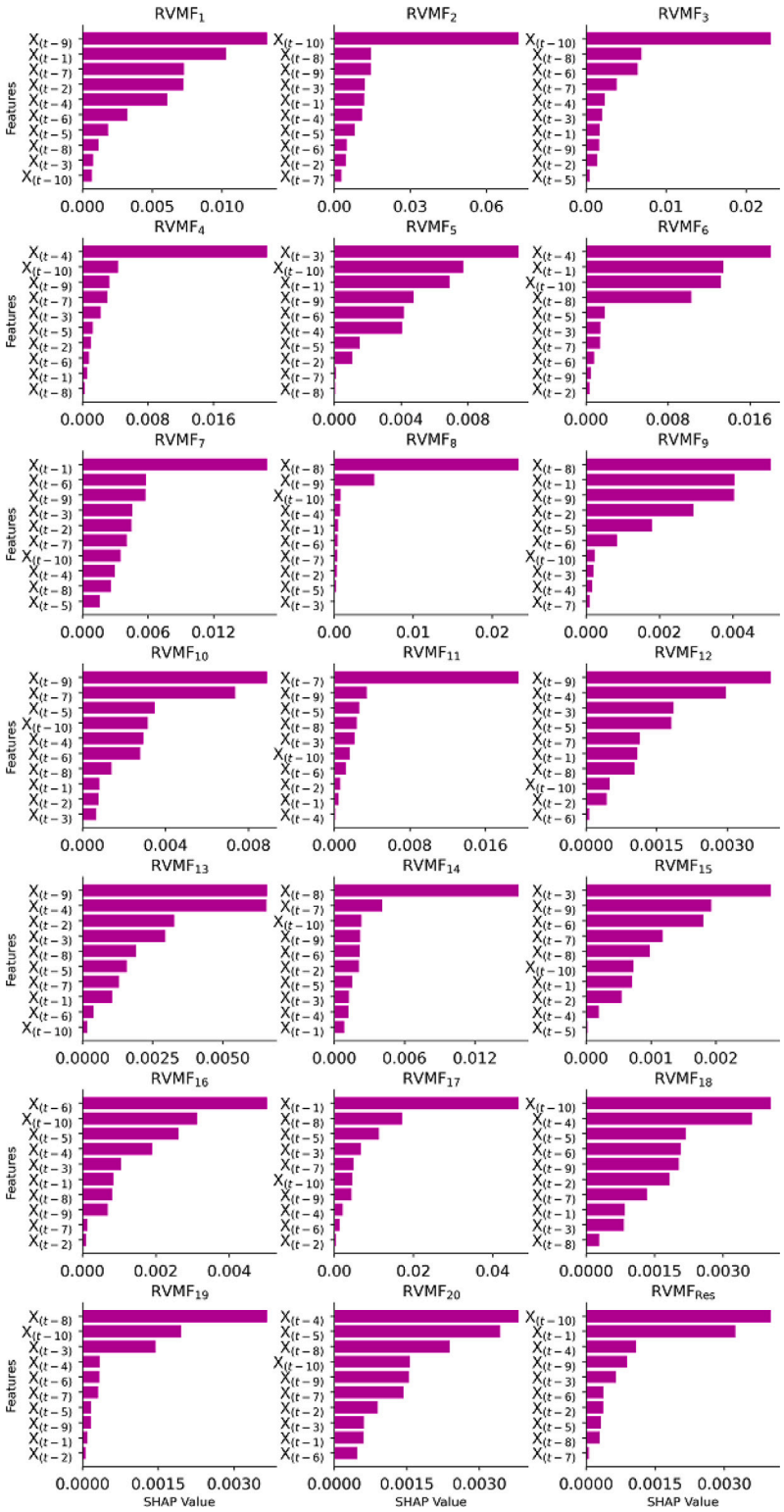


Fig. 13a. SHAP summary bar plots for the TAS. Note:- This plot is only for residual prediction using TabNet model after VMD decomposition. $RVMF$ is the residual variation mode function. $X_{(t-i)}$ where $i = 0, 1, 2, 3...10$, are the lagged values of $RVMF$ component.

- **SVMF$_{10}$**: The feature $X_{(t-8)}$ has the highest SHAP value, followed by $X_{(t-10)}$ and $X_{(t-4)}$, indicating their significant contribution to the model's predictions.
- **SVMF$_{13}$**: The feature $X_{(t-8)}$ has the highest SHAP value, followed by $X_{(t-10)}$ and $X_{(t-4)}$. This suggests that these features are crucial for the predictions made by the RFR model.
- **SVMF$_{19}$**: The feature $X_{(t-5)}$ has the highest SHAP value, indicating its significant impact on the model's predictions. Other features like $X_{(t-9)}$ and $X(t-1)$ also have notable contributions.
- **SVMF$_{Res}$**: This subplot shows that $X_{(t-1)}$ has the highest SHAP value, indicating it is the most influential feature for

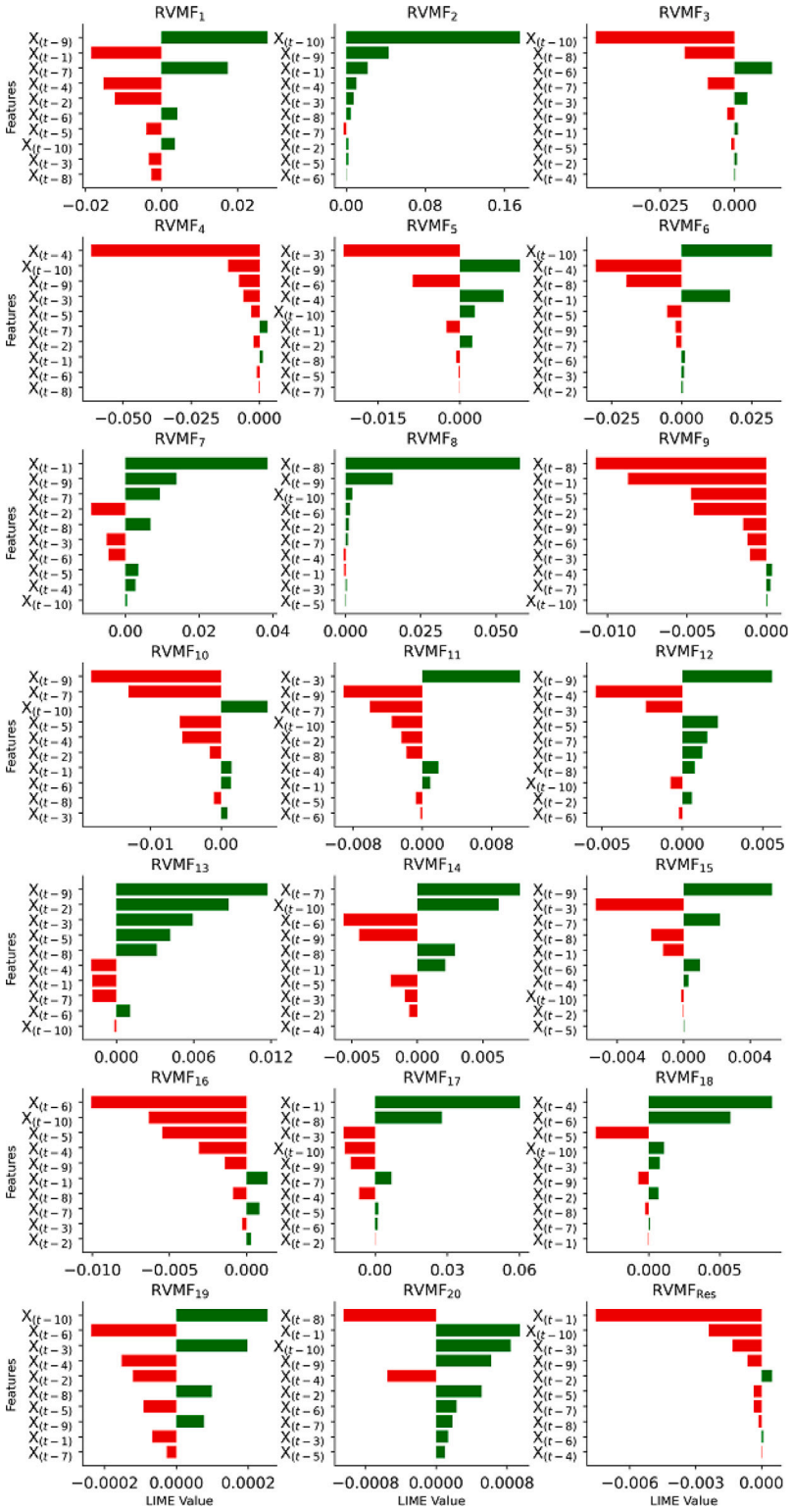


Fig. 13b. LIME explanation bar plots at instance 300 for TAS. Note:- This plot is only for residual prediction using TabNet model after VMD decomposition. $RVMF$ is the residual variation mode function.$X_{(t-i)}$ where $i = 0, 1, 2, 3...10$, are the lagged values of $RVMF$ component.

model prediction. Other features like $X_{(t-10)}$ and $X_{(t-8)}$ also contribute to the prediction.

Fig. 12b displays LIME plots for the RFR model used for seasonal prediction. Each subplot represents the prediction for a VMD sub-series using the RFR model for **instance 300**. The x-axis shows the LIME value, which indicates the importance of each feature in the model's prediction, while the y-axis lists the features, denoted as $X_{(t-n)}$.

(a) **SVMF$_1$**:

- **Top Features**: $X_{(t-1)}$, $X_{(t-2)}$, $X_{(t-3)}$.
- **LIME Values**: The features $X_{(t-1)}$, $X_{(t-2)}$, and $X_{(t-3)}$ have negative LIME values, indicating a negative contribution to the prediction.

(b) **SVMF$_3$**:

- **Top Features**: $X_{(t-8)}$, $X_{(t-3)}$, $X_{(t-9)}$.
- **LIME Values**: Mixed contributions, with $X_{(t-8)}$ and $X_{(t-3)}$ having negative LIME values and $X_{(t-9)}$ having a positive LIME value.

(c) **SVMF$_{10}$**:

- **Top Features**: $X_{(t-3)}$, $X_{(t-2)}$, $X_{(t-1)}$.
- **LIME Values**: All features have positive LIME values, indicating a positive contribution to the prediction.

(d) **SVMF$_{15}$**:

- **Top Features**: $X_{(t-5)}$, $X_{(t-3)}$, $X_{(t-9)}$.
- **LIME Values**: The feature $X_{(t-5)}$ has a positive LIME value, while $X_{(t-3)}$ and $X_{(t-9)}$ have negative LIME values.

(e) **SVMF$_{18}$**:

- **Top Features**: $X_{(t-8)}$, $X_{(t-9)}$, $X_{(t-10)}$.
- **LIME Values**: All features have positive LIME values, indicating a positive contribution to the prediction.

(f) **SVMF$_{Res}$**:

- **Top Features**: $X_{(t-1)}$.
- **LIME Values**: The feature $X_{(t-1)}$ has a negative LIME value, indicating a negative contribution to the prediction.

3. **Interpretability of TabNet model for residual component prediction:** The residual component of the STL for EP time-series is further decomposed utilizing VMD into 20 distinct variational mode functions, referred to as $RVMF$. Each resulting sub-series ($RVMF$) is subsequently predicted using the TabNet model. The input for the TabNet model comprises the lagged series, denoted as $X_{(t-n)}$, for each $RVMF$ component. Fig. 13a presents the SHAP value bar charts for multiple features across different RVMF sub-series. The RVMF$_{Res}$ represents the residual component following the VMD decomposition of the STL residual component. These SHAP value charts provide a visual representation of feature importance, where each subplot delineates the contribution of various features to the model's prediction. For instance, in the subplot for $RVMF_1$, the feature $X_{(t-9)}$ has the highest SHAP value, indicating it is the most influential feature for that particular model's predictions. Similarly, for $RVMF_2$, the feature $X_{(t-10)}$ stands out as the most significant contributor, followed by $X_{(t-8)}$ and $X_{(t-9)}$. This pattern of importance shifts across different $RVMF$ sub-series, demonstrating how each sub-series may rely on different features for accurate predictions. In another example, the subplot for $RVMF_7$ shows that $X_{(t-1)}$ and $X_{(t-6)}$ are the top contributing features. This variability highlights the model's dependence on recent and not-so-recent past data points for making predictions. Similarly, $RVMF_{14}$ indicates $X_{(t-8)}$ and $X_{(t-7)}$ as the most significant features, suggesting a strong influence from the earlier data points. The residual model $RVMF_{Res}$ also follows a similar pattern, where $X_{(t-10)}$ and $X_{(t-1)}$ have the highest SHAP values, indicating their substantial contribution to the residual predictions.

Similarly, Fig. 13b illustrates the LIME plots. Each subplot provides a detailed visualization of the contributions of various features to the predictions made by a TabNet model for **instance 300**. The x-axis in these subplots represents the LIME values, which indicate the importance and direction (positive or negative) of each feature's contribution to the model's prediction. The y-axis lists the features as $X_{(t-1)}, X_{(t-2)}, \ldots, X_{(t-10)}$, showing the lagged time-series data points used in the TabNet model for STL decomposed residual prediction. In the analysis of specific $RVMF$ sub-series, several key observations can be made. For $RVMF_1$, the feature $X_{(t-9)}$ has a significant positive LIME value, indicating a strong positive contribution to the prediction, while $X_{(t-1)}$ has a negative LIME value, contributing negatively to the prediction. In $RVMF_2$, the feature $X_{(t-10)}$ has a substantial positive LIME value, making it the most influential feature for the model's prediction in this instance, with $X_{(t-9)}$ and $X_{(t-1)}$ also contributing positively but to a lesser extent. For $RVMF_5$, $X_{(t-3)}$ shows the highest negative LIME value, indicating its strong negative impact on the prediction, whereas $X_{(t-9)}$ has a positive LIME value, suggesting a positive contribution. The subplot for $RVMF_{10}$ reveals that $X_{(t-9)}$ has the highest negative LIME value, significantly lowering the prediction, while $X_{(t-10)}$ makes a notable positive contribution. In $RVMF_{14}$, $X_{(t-7)}$ stands out with a strong positive LIME value, making it the most influential feature for the prediction, while $X_{(t-6)}$ and $X_{(t-9)}$ have negative LIME values, indicating they negatively impact the prediction. Furthermore, for $RVMF_{19}$, the features $X_{(t-10)}$ and $X_{(t-3)}$ exhibit positive LIME values, contributing positively to the prediction, whereas $X_{(t-6)}$ shows a negative LIME value, suggesting a negative impact. Lastly, for $RVMF_{Res}$,all features have negative contribution to the prediction except the $X_{(t-2)}$ and $X_{(t-6)}$. This variation in feature contributions underscores the complexity of the model's behavior and the importance of understanding how each feature impacts different components of the prediction using LIME approach.

In summary, utilizing visual representations with SHAP and LIME facilitates the understanding of which features exert the most influence on predictions generated by D3Net model. These plots offer clear global and local explanations of feature importance, significantly enhancing the interpretability and transparency of proposed D3Net model in half-hourly EP prediction.

Given the results presented, the need for explainability in AI-based models has become essential across various fields due to regulatory demands for capabilities like traceability, transparency, and interpretability of models [88,89]. To foster trustworthy artificial intelligence in the domain of EP forecasting, it is vital to ensure that artificial intelligence models comply with legal and ethical standards [89]. Moreover, the robust AI-based deep learning systems are highly sensitive to minor changes in practice, which complicates the explanation of the forecasted results. Perturbations in input data can significantly affect results, especially since many critical areas suffer from low-quality datasets that do not meet the standards of independent and identically distributed (i.i.d.) data [88]. Thus, our research addresses these critical issues, by exploiting advanced model-agnostic *xAI* tools to implement trustworthy artificial intelligence in the EP forecasting system.

The newly developed hybridized explainable D3Net model based on *xAI* approach in this study can facilitate the energy experts and grid operators in decision-making by providing consistent, transparent and more accurate forecasts of EP. Hence, the trustworthiness and performance superiority of this smart predictive framework extends its practical utility as a decision support tool to enhance the overall grid reliability, cost-effectiveness and efficiency for the energy sector in the near future.

Table 11
Average of computation time.

Model	Construction time (Training and Validation, Minute)	Testing (Sec)
D3Net	152	109
STL-MLP	67	87
STL-RF	52	69
STL-CNN	66	72
STL-TabNet	89	87
STL-CLSTM	123	100
STL-LSTM	134	100
TabNet	97	96
STL-DNN	137	99
MLP	53	63
RFR	45	55
LSTM	67	65
CLSTM	76	43
DNN	87	87
CNN	91	65

4.2. Computational resource requirements

The computational time of a prediction model is critical for utility companies, particularly in scenarios involving online training. In such cases, daily electricity price observations are continually incorporated into the training dataset for model retraining, making computational time a key factor. The time required for electricity price prediction is influenced by factors such as the length of the moving window, the number of predictors, and, most significantly, the choice of prediction model. Table 11 compares the computation times of the proposed D3Net model with fourteen benchmark models. The results indicate that the D3Net model, while less computationally efficient than some of the other models, remains operational for an extended period once trained. Furthermore, the testing time is under two minute, making the proposed D3Net model suitable for practical applications in real-time electricity price forecasting. The simulations were performed on an Intel® Core™ i9 10th Generation processor, operating at 3.8 GHz with 32 GB of memory. Despite its relatively higher training time, the D3Net model's performance in terms of accuracy and prediction quality makes it a viable choice for electricity price forecasting in utility operations, where the need for precision outweighs computational efficiency in the context of retraining and deployment over time.

Conclusion and future work

Accurate prediction of electricity prices (EP) empowers utilities and grid operators to make well-informed decisions regarding energy trading, grid stability, resource allocation, and pricing strategies. This enhances overall grid reliability, efficiency, and cost-effectiveness. This study introduces a novel D3Net model that integrates STL and VMD with MLP, RFR, and TabNet as prediction models specifically designed for half-hourly EP prediction. The methodology comprises several different steps, including extracting trend of EP time-series and its prediction with an MLP model. The seasonal component is further decomposed using VMD into 20 VMFs to capture underlying seasonal patterns more effectively. Then, each VMF is predicted using a RFR model. The residual component from STL decomposition undergoes another round of VMD decomposition, with each resulting subseries predicted using the TabNet model. Finally, predictions from the trend, seasonal, and residual components are combined to generate the final prediction.

Additionally, fourteen standalone and STL decomposition-based models, including LSTM, Convolutional LSTM, CNN, RFR, MLP, DNN, and variations such as STL-LSTM, STL-CLSTM, STL-CNN, STL-RFR, STL-MLP, STL-DNN, and STL-TabNet, were developed for comparative analysis. The effectiveness of the proposed D3Net model was assessed using a range of deterministic metrics, statistical metrics and diagnostic plots. Historical data spanning from 2019 to 2022 were employed to evaluate the model's accuracy in predicting half-hourly EP for SA and TAS.

It can be concluded that:

- The superiority of the proposed D3Net model is evident with low RMSE and MAE metrics: RMSE of $\approx$ 4.629 AUD/MWh and MAE of $\approx$ 3.051 AUD/MWh for SA, and RMSE of $\approx$ 4.133 AUD/MWh and MAE of $\approx$ 2.562 AUD/MWh for TAS. In contrast, the standalone CNN model performed significantly worse, with RMSE of $\approx$ 29.46 AUD/MWh and MAE of 14.26 AUD/MWh for SA, and RMSE of $\approx$ 29.45 AUD/MWh and MAE of $\approx$ 13.158 AUD/MWh for TAS during testing.
- In terms of relative error, WI, NS and LM, the D3Net reaches highest level of precision [$RRMSE \approx 7.9\%$, $RMAE \approx 5.2\%$, $WI \approx 0.998$, $NS \approx 0.991$, $LM \approx 0.961$] for SA; and [$RRMSE \approx 5.4\%$, $RMAE \approx 3.4\%$, $WI \approx 0.996$, $NS \approx 0.985$, $LM \approx 0.946$] for TAS in predicting half-hourly EP.
- Based on GPI metrics, which consolidate all analyzed models into a single parameter, models can be easily ranked to assist in selecting the most appropriate and accurate ones. According to GPI values, the D3Net model demonstrates superiority with GPI scores of $\approx$ 11.068 for SA and $\approx$ 12.206 for TAS.
- Statistical tests including the Diebold–Mariano (Fig. 10) and Giacomini–White tests (Fig. B.2a and Fig. B.2b) demonstrated that predictions from the D3Net models exhibit significantly higher accuracy compared to those from the benchmark models.
- A Wilcoxon signed-rank test (Table 10) was also conducted to further evaluate the proposed D3Net model against other standalone and decomposition-based models, assessing the significance of its double decomposition approach. This test confirms that the D3Net model significantly outperforms the benchmark models, reinforcing the novelty and effectiveness of the proposed algorithm.
- Finally, the SHAP interpretability components provide valuable insights into how individual features contribute to D3Net model predictions (refer Figs. 11a, 12a and 13a) and the LIME provides insights into local instances by explaining individual predictions (refer Figs. 11b, 12b and 13b) generated by D3Net models.

Beyond forecasting EP, the exceptional performance and reliability of the D3Net model indicate its potential usefulness to a wide range of alternative time-series problems. Future research will focus on evaluating the model's effectiveness in incorporating factors such as electricity demand, electricity production, solar and wind generation, and climatological variables including air temperature, solar radiation, rainfall, relative humidity, and evaporation, among others. The *xAI* approach proposed in this study can also improved with additional inputs for price prediction such as, but not limited to regional weather data, economy growth data, population growth data, etc. to further

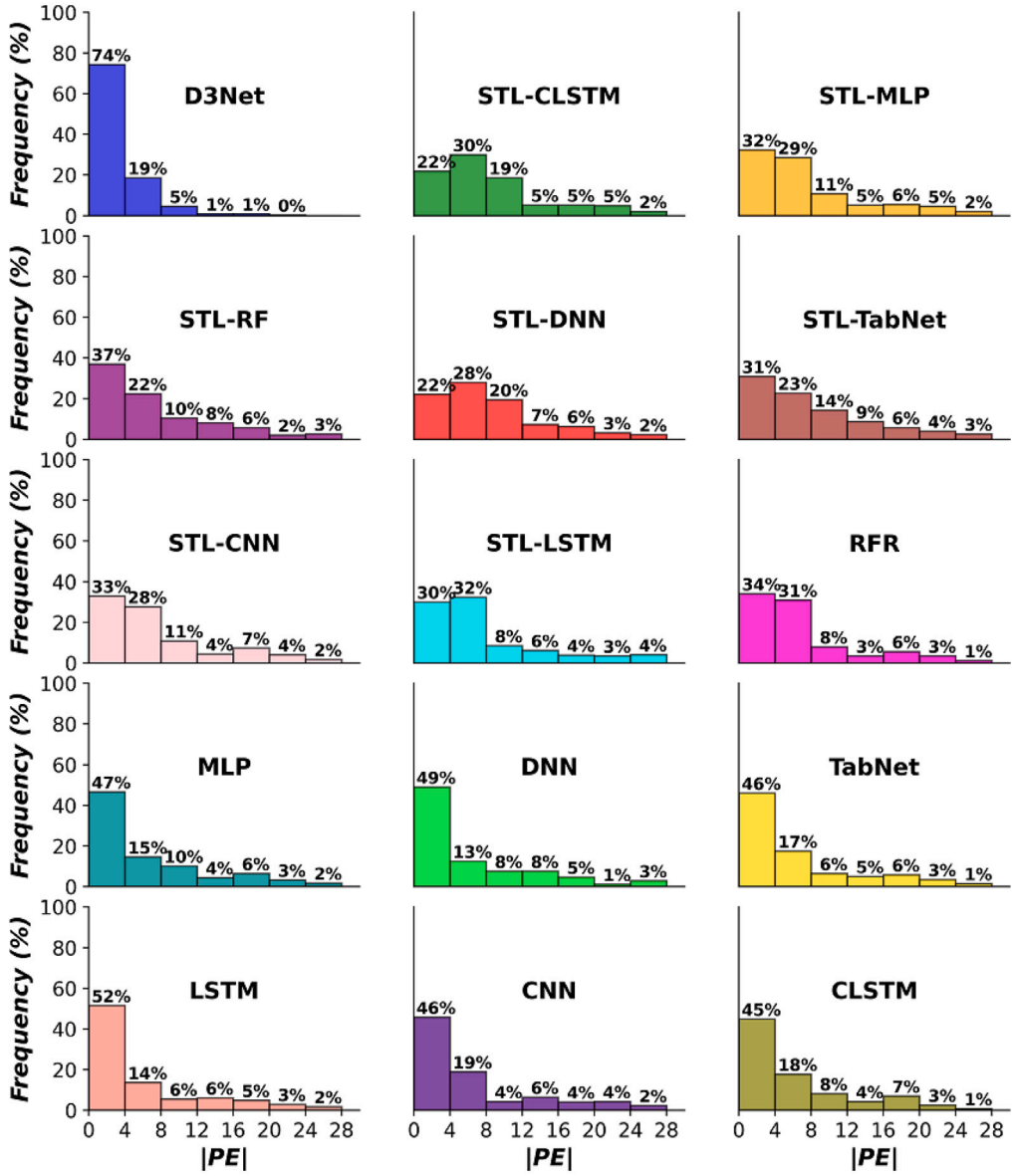


Fig. B.1a. Histogram of the frequency distribution of the absolute Prediction Error (PE) calculated for the test period for SA. The numbers on each error bar show cumulative percentage frequency of the occurrence of the respective error represented by that particular bar.

research ascertain the robustness and practical use of the proposed model.

CRediT authorship contribution statement

Sujan Ghimire: Writing – original draft, Software, Investigation, Data curation, Conceptualization. **Ravinesh C. Deo:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Konstantin Hopf:** Writing – review & editing, Validation, Supervision, Investigation, Conceptualization. **Hangyue Liu:** Writing – review & editing, Validation, Supervision, Investigation. **David Casillas-Pérez:** Writing – review & editing, Visualization, Validation, Investigation. **Andreas Helwig:** Writing – review & editing, Validation, Supervision. **Salvin S. Prasad:** Writing – review & editing, Visualization, Investigation. **Jorge Pérez-Aracil:** Writing – review & editing, Validation, Investigation. **Prabal Datta Barua:** Writing – review & editing. **Sancho Salcedo-Sanz:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used generative AI to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank data providers, all reviewers and Editor for their thoughtful comments, suggestions and review process. Partial support of this work through projects PID2023-150663NB-C21 and PID2022-140786NB-C31 (LATENTIA) of the Spanish Ministry of Science, Innovation and Universities (MICINNU), is also acknowledged.

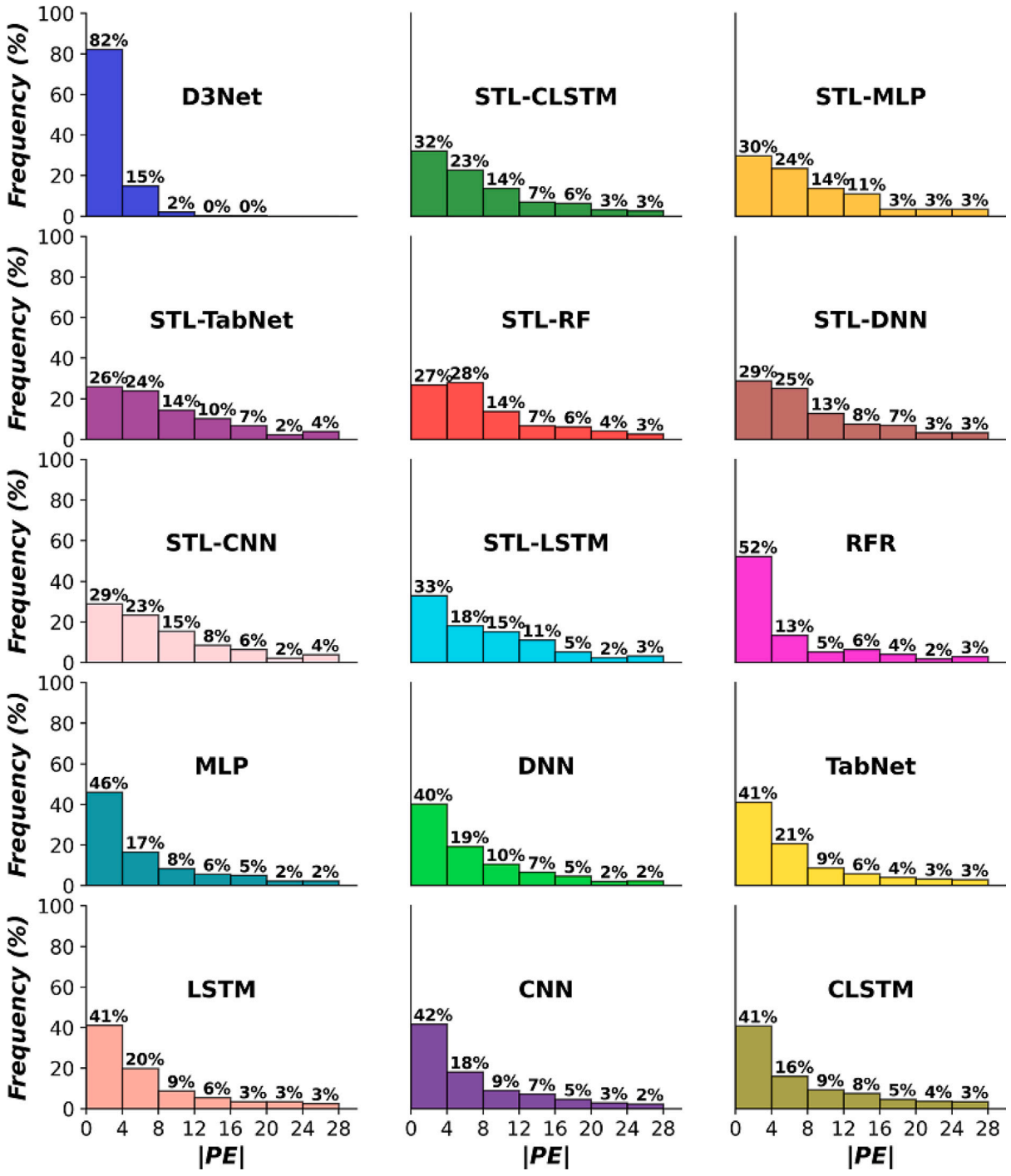


Fig. B.1b. Same as Fig. B.1a for TAS.

Appendix A. Acronyms

ANN	Artificial Neural Network
AEMO	Australian Energy Market Operator
AUD	Australian Dollar
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARMA	Autoregressive Moving Average
CNN	Convolutional Neural Network
CLSTM	Convolutional LSTM
DL	Deep Learning
DT	Decision Tree
DNN	Deep Neural Network
EMD	Empirical Mode Decomposition
GRU	Gated Recurrent Unit
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
LIME	Local Interpretable Model-Agnostic Explanations
LSTM	Long Short-Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
PACF	Partial Autocorrelation Function
PSO	Particle Swarm Optimization

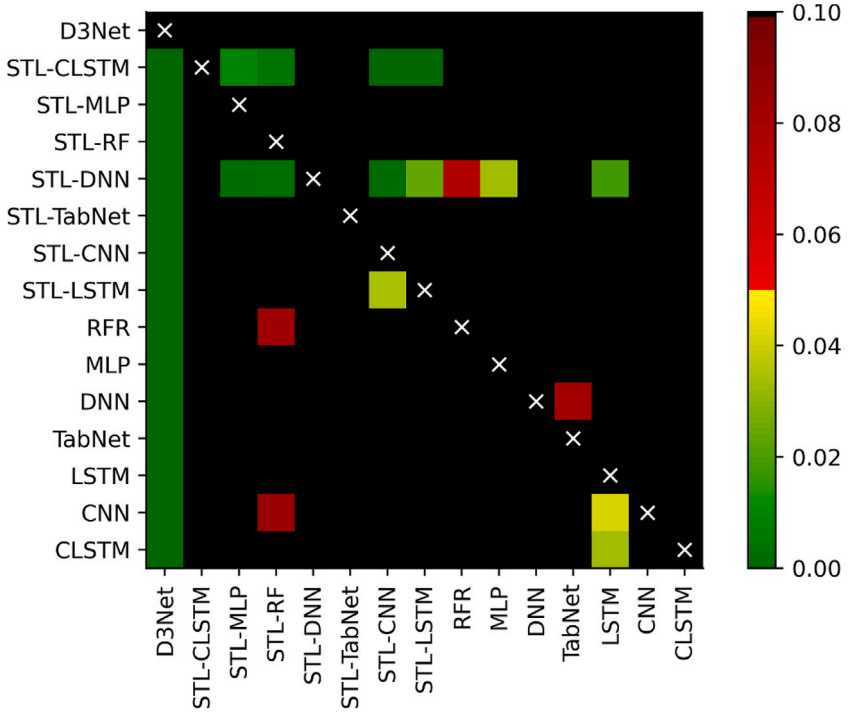


Fig. B.2a. Comparison of prediction outcomes using statistical *GW* test for SA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

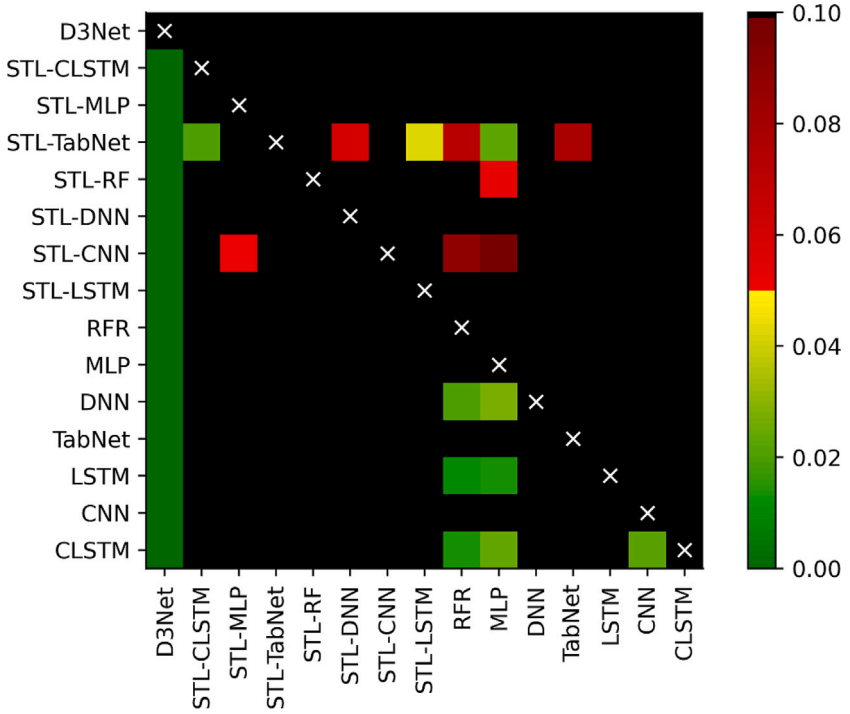


Fig. B.2b. Comparison of prediction outcomes using statistical *GW* test for TAS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

RFR Random Forest Regression
RNN Recurrent Neural Network
RMSE Root Mean Square Error
SA South Australia
SDG Sustainable Development Goal
SHAP SHapley Additive exPlanations
STL Seasonal-Trend decomposition using LOESS
SVR Support Vector Regression
TAS Tasmania
VMD Variational Mode Decomposition
VMF Variational Mode Function
VAR Vector Autoregressive Model

Appendix B. Other experiments

B.1. Prediction error frequency

To assess the frequency of different Prediction Errors (PE) during the testing period, a histogram was constructed showing the distribution of absolute PE ($|PE|$)in error bins spanning ± 4 AUD/MWh (see Figs. B.1a and B.1b). For each error bin, the total number of instances (or the percentage of occurrences within that error band) is plotted on the ordinate. Notably, for the smallest error magnitudes (± 4 AUD/MWh) considered, the D3Net model shows the highest frequency, with $\approx$ 74% for SA and $\approx$ 82% for TAS. In contrast, the STL-CLSTM, STL-MLP, STL-TabNet, STL-RF, STL-DNN,STL-CNN, STL-LSTM, RFR, MLP, DNN, TabNet, LSTM, CNN and CLSTM models record frequencies of $\approx$ 22%, $\approx$ 32%, $\approx$ 37%, $\approx$ 22%, $\approx$ 31%, $\approx$ 33%, $\approx$ 30%, $\approx$ 34%, $\approx$ 47%, $\approx$ 49%, $\approx$ 46%, $\approx$ 52%, $\approx$ 46%, and $\approx$ 45% respectively, for SA. This analysis revealed that the D3Net model consistently reduced errors more effectively than the counterpart models, demonstrating significant improvements in electricity price (EP) prediction.

B.2. The statistical Giacomini–White test

Similarly, Figs. B.2a and B.2b display the range of $p-values$ represented by a heatmap utilizing GW statistics. In this heatmap, dark green indicates $p-values$ close to zero, signifying a significant difference in prediction performance between the models compared on the $X-axis$ and $Y-axis$. Conversely, $p-values$ equal to or greater than 0.10 are represented in black. The heatmap clearly shows that the D3Net model significantly outperforms all other competing models, as indicated by the dark green cells corresponding to D3Net's comparisons with other models. Therefore, the performance statistics using GW statistics align with previous findings, reaffirming that D3Net is the best-performing model.

Data availability

Data were acquired from Australian Energy Market Operator (AEMO). https://www.aemo.com.au/.

References

[1] Gürtler M, Paulsen T. Forecasting performance of time series models on electricity spot markets: a quasi-meta-analysis. Int J Energy Sect Manag 2018;12(1):103–29.

[2] Khan AM, Osińska M. Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India. Expert Syst Appl 2023;212:118840.

[3] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Appl Energy 2018;221:386–405.

[4] Valente JM, Maldonado S. SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression. Expert Syst Appl 2020;160:113729.

[5] Atef S, Eltawil AB. A comparative study using deep learning and support vector regression for electricity price forecasting in smart grids. In: 2019 IEEE 6th international conference on industrial engineering and applications. ICIEA, IEEE; 2019, p. 603–7.

[6] Fragkioudaki A, Marinakis A, Cherkaoui R. Forecasting price spikes in European day-ahead electricity markets using decision trees. In: 2015 12th international conference on the European energy market. EEM, IEEE; 2015, p. 1–5.

[7] Mei J, He D, Harley R, Habetler T, Qu G. A random forest method for real-time price forecasting in new york electricity market. In: 2014 IEEE PES general meeting| conference & exposition. IEEE; 2014, p. 1–5.

[8] Wu K, Chai Y, Zhang X, Zhao X. Research on power price forecasting based on pso-xgboost. Electronics 2022;11(22):3763.

[9] Sansom DC, Downs T, Saha TK. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. J Electr Electron Eng Aust 2003;22(3):227–33.

[10] Azmira WAI, Ahmad A, Abidin IZ, Yap KS, Nasir MNM, Upkli WR. Electricity price prediction with support vector machine and bacterial foraging optimization algorithm for day-ahead model. In: 2020 IEEE student conference on research and development. sCOReD, IEEE; 2020, p. 159–64.

[11] Panapakidis IP, Dagoumas AS. Day-ahead electricity price forecasting via the application of artificial neural network based models. Appl Energy 2016;172:132–51.

[12] Pourdaryaei A, Mokhlis H, Illias HA, Kaboli SHA, Ahmad S, Ang SP. Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in de-regulated electricity market. Ieee Access 2019;7:125369–86.

[13] Dai X, Liu G-P, Hu W. An online-learning-enabled self-attention-based model for ultra-short-term wind power forecasting. Energy 2023;272:127173.

[14] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. 2014, arXiv preprint arXiv:1412.3555.

[15] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.

[16] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85.

[17] Ugurlu U, Oksuz I, Tas O. Electricity price forecasting using recurrent neural networks. Energies 2018;11(5):1255.

[18] Xiong X, Qing G. A hybrid day-ahead electricity price forecasting framework based on time series. Energy 2023;264:126099.

[19] Kapoor G, Wichitaksorn N. Electricity price forecasting in New Zealand: A comparative analysis of statistical and machine learning models with feature selection. Appl Energy 2023;347:121446.

[20] Sezer OB, Gudelek MU, Ozbayoglu AM. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Appl Soft Comput 2020;90:106181.

[21] Deng Z, Liu C, Zhu Z. Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network. Int J Electr Power Energy Syst 2021;125:106499.

[22] Wang Y, Wang J, Li Z. A novel hybrid air quality early-warning system based on phase-space reconstruction and multi-objective optimization: A case study in China. J Clean Prod 2020;260:121027.

[23] Heidarpanah M, Hooshyaripor F, Fazeli M. Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market. Energy 2023;263:126011.

[24] Das R, Bo R, Chen H, Rehman WU, Wunsch D. Forecasting nodal price difference between day-ahead and real-time electricity markets using long-short term memory and sequence-to-sequence networks. IEEE Access 2021;10:832–43.

[25] Li Y, Ding Y, Liu Y, Yang T, Wang P, Wang J, et al. Dense skip attention based deep learning for day-ahead electricity price forecasting. IEEE Trans Power Syst 2022;38(5):4308–27.

[26] Li G, Li F, Ahmad T, Liu J, Li T, Fang X, et al. Performance evaluation of sequence-to-sequence-attention model for short-term multi-step ahead building energy predictions. Energy 2022;259:124915.

[27] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. Appl Energy 2024;353:122059.

[28] Pourdaryaei A, Mohammadi M, Mubarak H, Abdellatif A, Karimi M, Gryazina E, et al. A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market. Expert Syst Appl 2024;235:121207.

[29] O'Connor C, Collins J, Prestwich S, Visentin A. Electricity price forecasting in the irish balancing market. Energy Strat Rev 2024;54:101436.

[30] Laitsos V, Vontzos G, Bargiotas D, Daskalopulu A, Tsoukalas LH. Data-driven techniques for short-term electricity price forecasting through novel deep learning approaches with attention mechanisms. Energies 2024;17(7):1625.

[31] Ehsani B, Pineau P-O, Charlin L. Price forecasting in the ontario electricity market via TriConvGRU hybrid model: Univariate vs. multivariate frameworks. Appl Energy 2024;359:122649.

[32] Monjazeb MR, Amiri H, Movahedi A. Wholesale electricity price forecasting by quantile regression and Kalman filter method. Energy 2024;290:129925.

[33] Xu Y, Li J, Wang H, Du P. A novel probabilistic forecasting system based on quantile combination in electricity price. Comput Ind Eng 2024;187:109834.

[34] Kuo P-H, Huang C-J. An electricity price forecasting model by hybrid structured deep neural networks. Sustainability 2018;10(4):1280.

[35] Lehna M, Scheller F, Herwartz H. Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account. Energy Econ 2022;106:105742.

[36] Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. Energy 2019;187:115804.

[37] Qiao W, Yang Z. Forecast the electricity price of US using a wavelet transform-based hybrid model. Energy 2020;193:116704.

[38] Huang C-J, Shen Y, Chen Y-H, Chen H-C. A novel hybrid deep neural network model for short-term electricity price forecasting. Int J Energy Res 2021;45(2):2511–32.

[39] Alkawaz AN, Abdellatif A, Kanesan J, Khairuddin ASM, Gheni HM. Day-ahead electricity price forecasting based on hybrid regression model. IEEE Access 2022;10:108021–33.

[40] Alamaniotis M, Bargiotas D, Bourbakis NG, Tsoukalas LH. Genetic optimal regression of relevance vector machines for electricity pricing signal forecasting in smart grids. IEEE Trans Smart Grid 2015;6(6):2997–3005.

[41] Zhou S, Zhou L, Mao M, Tai H-M, Wan Y. An optimized heterogeneous structure LSTM network for electricity price forecasting. Ieee Access 2019;7:108161–73.

[42] Torres JF, Martínez-Álvarez F, Troncoso A. A deep LSTM network for the spanish electricity consumption forecasting. Neural Comput Appl 2022;34(13):10533–45.

[43] Zhang T, Tang Z, Wu J, Du X, Chen K. Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. Electr Power Syst Res 2022;205:107762.

[44] Pourdaryaei A, Mokhlis H, Illias HA, Kaboli SHA, Ahmad S. Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach. Ieee Access 2019;7:77674–91.

[45] Yang W, Wang J, Niu T, Du P. A novel system for multi-step electricity price forecasting for electricity market management. Appl Soft Comput 2020;88:106029.

[46] Yan L, Yan Z, Li Z, Ma N, Li R, Qin J. Electricity market price prediction based on quadratic hybrid decomposition and THPO algorithm. Energies 2023;16(13):5098.

[47] He K, Yu L, Tang L. Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology. Energy 2015;91:601–9.

[48] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for short term electricity price forecasting. Appl Energy 2020;258:114087.

[49] Guo F, Deng S, Zheng W, Wen A, Du J, Huang G, et al. Short-term electricity price forecasting based on the two-layer VMD decomposition technique and SSA-LSTM. Energies 2022;15(22):8445.

[50] Zhang B, Song C, Jiang X, Li Y. Electricity price forecast based on the STL-TCN-NBEATS model. Heliyon 2023;9(1).

[51] Ramírez-García A, Saucedo E. Hedging electricity price volatility applying seasonal and trend decomposition. Análisis Económico 2022;37(94):143–66.

[52] Ding W, Abdel-Basset M, Hawash H, Ali AM. Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey. Inform Sci 2022;615:238–92.

[53] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.

[54] Ribeiro MT, Singh S, Guestrin C. " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, p. 1135–44.

[55] Xu L, Ou Y, Cai J, Wang J, Fu Y, Bian X. Offshore wind speed assessment with statistical and attention-based neural network methods based on STL decomposition. Renew Energy 2023;216:119097.

[56] Stefenon SF, Seman LO, Mariani VC, Coelho LdS. Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices. Energies 2023;16(3):1371.

[57] Li W, Jiang X. Prediction of air pollutant concentrations based on TCN-BiLSTM-DMAttention with STL decomposition. Sci Rep 2023;13(1):4665.

[58] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2013;62(3):531–44.

[59] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Electricity demand error corrections with attention bi-directional neural networks. Energy 2024;291:129938.

[60] Breiman L. Random forests. Mach Learn 2001;45:5–32.

[61] Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.. Psychol Methods 2009;14(4):323.

[62] Joelsson SR, Benediktsson JA, Sveinsson JR. Feature selection for morphological feature extraction using randomforests. In: Proceedings of the 7th nordic signal processing symposium-NORSIG 2006. IEEE; 2006, p. 138–41.

[63] Strobl C, Boulesteix A-L, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. BMC Bioinformatics 2008;9:1–11.

[64] Deo RC, Ghimire S, Downs NJ, Raj N. Optimization of windspeed prediction using an artificial neural network compared with a genetic programming model. In: Handbook of research on predictive modeling and optimization methods in science and engineering. IGI Global; 2018, p. 328–59.

[65] Chen S-A, Li C-L, Yoder N, Arik SO, Pfister T. Tsmixer: An all-mlp architecture for time series forecasting. 2023, arXiv preprint arXiv:2303.06053.

[66] Liu J, Huang X, Li Q, Chen Z, Liu G, Tai Y. Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. Energy Convers Manage 2023;280:116804.

[67] Arik SÖ, Pfister T. Tabnet: Attentive interpretable tabular learning. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, (8):2021, p. 6679–87.

[68] Wang Y, Abliz A, Ma H, Liu L, Kurban A, Halik Ü, et al. Hyperspectral estimation of soil copper concentration based on improved TabNet model in the eastern junggar coalfield. IEEE Trans Geosci Remote Sens 2022;60:1–20.

[69] Nguyen HV, Byeon H. Prediction of out-of-hospital cardiac arrest survival outcomes using a hybrid agnostic explanation TabNet model. Mathematics 2023;11(9):2030.

[70] Deo RC, Grant RH, Webb A, Ghimire S, Igoe DP, Downs NJ, et al. Forecasting solar photosynthetic photon flux density under cloud cover effects: novel predictive model using convolutional neural network integrated with long short-term memory network. Stoch Environ Res Risk Assess 2022;36(10):3183–220.

[71] Ghimire S, Nguyen-Huy T, AL-Musaylh MS, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Integrated multi-head self-attention transformer model for electricity demand prediction incorporating local climate variables. Energy AI 2023;14:100302.

[72] Marinho FP, Rocha PA, Neto AR, Bezerra FD. Short-term solar irradiance forecasting using CNN-1D, LSTM, and CNN-LSTM deep neural networks: A case study with the Folsom (USA) dataset. J Sol Energy Eng 2023;145(4):041002.

[73] Namdari H, Haghighi A, Ashrafi SM. Short-term urban water demand forecasting; application of 1D convolutional neural network (1D CNN) in comparison with different deep learning schemes. Stoch Environ Res Risk Assess 2023;1–16.

[74] Jayasinghe WLP, Deo RC, Ghahramani A, Ghimire S, Raj N. Development and evaluation of hybrid deep learning long short-term memory network model for pan evaporation estimation trained with satellite and ground-based data. J Hydrol 2022;607:127534.

[75] Memarzadeh G, Keynia F. A new hybrid CBSA-GA optimization method and MRMI-LSTM forecasting algorithm for PV-ESS planning in distribution networks. J Energy Storage 2023;72:108582.

[76] Elbeltagi A, Srivastava A, Li P, Jiang J, Jinsong D, Rajput J, et al. Forecasting actual evapotranspiration without climate data based on stacked integration of DNN and meta-heuristic models across China from 1958 to 2021. J Environ Manag 2023;345:118697.

[77] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach. Energy Convers Manage 2023;297:117707.

[78] Ghimire S, Deo RC, Pourmousavi SA, Casillas-Pérez D, Salcedo-Sanz S. Point-based and probabilistic electricity demand prediction with a neural facebook prophet and kernel density estimation model. Eng Appl Artif Intell 2024;135:108702.

[79] Nanjappa Y, Nassa VK, Varshney G, Lal B, Pandey S, Turukmane AV. Improving migration forecasting for transitory foreign tourists using an ensemble DNN-LSTM model. Entertain Comput 2024;50:100665.

[80] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, Pourmousavi SA, Acharya UR. Probabilistic-based electricity demand forecasting with hybrid convolutional neural network-extreme learning machine model. Eng Appl Artif Intell 2024;132:107918.

[81] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. 2019, p. 2623–31.

[82] Srinivas P, Katarya R. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. Biomed Signal Process Control 2022;73:103456.

[83] Hanifi S, Lotfian S, Zare-Behtash H, Cammarano A. Offshore wind power forecasting—A new hyperparameter optimisation algorithm for deep learning models. Energies 2022;15(19):6919.

[84] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Review and statistical analysis of different global solar radiation sunshine models. Renew Sustain Energy Rev 2015;52:1869–80.

[85] Diebold FX. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. J Bus Econom Statist 2015;33(1). 1–1.

[86] Giacomini R, White H. Tests of conditional predictive ability. Econometrica 2006;74(6):1545–78.

[87] Rosner B, Glynn RJ, Lee M-LT. The wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006;62(1):185–92.

[88] Holzinger A. The next frontier: AI we can really trust. In: Joint European conference on machine learning and knowledge discovery in databases. Springer; 2021, p. 427–40.

[89] Holzinger A, Dehmer M, Emmert-Streib F, Cucchiara R, Augenstein I, Del Ser J, et al. Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence. Inf Fusion 2022;79:263–78.