Applied Energy 374 (2024) 123920

Contents lists available at ScienceDirect

Applied Energy

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

Half-hourly electricity price prediction with a hybrid convolution neural network-random vector functional link deep learning approach

Sujan Ghimire [a], Ravinesh C. Deo [a,*], David Casillas-Pérez [b], Ekta Sharma [a], Sancho Salcedo-Sanz [c,a], Prabal Datta Barua [d,e,f], U. Rajendra Acharya [a]

[a] *Artificial Intelligence Applications Laboratory, School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD, 4300, Australia*
[b] *Department of Signal Processing and Communications, Universidad Rey Juan Carlos, Fuenlabrada, 28942, Madrid, Spain*
[c] *Department of Signal Processing and Communications, Universidad de Alcalá, Alcalá de Henares, 28805, Madrid, Spain*
[d] *Cogninet AI, Sydney, NSW 2010, Australia*
[e] *School of Business (Information System), University of Southern Queensland, Australia*
[f] *Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia*

ARTICLE INFO

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

MSC:
0000
1111

Keywords:
Deep learning
Convolutional neural network
Maximum overlap discrete wavelet transform
Hybrid models
Random vector functional link

ABSTRACT

Digital technologies with predictive modelling capabilities are revolutionizing electricity markets, especially in demand-side management. Accurate electricity price prediction is essential in deregulated markets; however, developing effective models is challenging due to high-frequency fluctuations and price volatility. This study introduces a hybrid prediction system that addresses these challenges through a comprehensive data processing and modelling framework for half-hourly electricity price predictions. The preprocessing stage employs the Maximum Overlap Discrete Wavelet Transform (MoDWT) to enhance input quality by reducing overlap and revealing underlying price patterns. The prediction model integrates Convolutional Neural Networks with Random Vector Functional Link (CRVFL) in a deep learning hybrid approach. Bayesian Optimization fine-tunes the MoDWT-CRVFL model for optimal performance. Validation of the model is conducted using half-hourly electricity prices from New South Wales. The results highlight the efficacy of the MoDWT-CRVFL model, achieving high accuracy with superior Global Performance Indicator (GPI) values of approximately 1.61, 1.33, 1.85, 1.30, and 0.78 for Summer, Autumn, Winter, Spring, and Annual (Year 2022), respectively, outperforming alternative models. Similarly, the Kling–Gupta Efficiency (KGE) metrics for the proposed model consistently surpassed those of both decomposition-based and standalone models. For instance, the KGE value for MoDWT-CRVFL was approximately 0.972, significantly higher than values of approximately 0.958, 0.899, 0.963, 0.943, 0.930, 0.661, 0.708, 0.696, 0.739, and 0.738 for MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, MoDWT-MLP, Bi-LSTM, LSTM, DNN, RF, XGB, and MLP, respectively. The methodologies proposed in this study optimize energy resource allocation, market prices, and network management, empowering market operators to make informed decisions for a resilient and efficient electricity market.

1. Introduction

Due to the constantly changing mechanisms and intense competition in the power industry, predicting electricity pricing has become a vital research field [1]. Prediction methods and models are widely used and highly regarded in the industry. Accurate and efficient prediction of electricity prices (*EP*) is crucial for energy-market players to make informed decisions regarding risk management in competitive electricity markets and optimize their profits [2]. Market players use price prediction to determine their bidding strategies, allocate resources, and plan for facility investment. The primary objective of electricity market players is to establish a transparent and efficient market. Achieving this goal needs the accurate and dependable prediction of future *EP*, which is crucial in formulating bidding strategies and maximizing profits [3]. However, the *EP* dataset is characterized by frequent fluctuations, non linearity, non-stationarity, and randomness, making it difficult to forecast electricity prices accurately. Additionally, various factors

---

[*] Corresponding author.
E-mail addresses: sujan.ghimire@unisq.edu.au (S. Ghimire), ravinesh.deo@unisq.edu.au (R.C. Deo), david.casillas@urjc.es (D. Casillas-Pérez), ekta.sharma@unisq.edu.au (E. Sharma), sancho.salcedo@uah.es (S. Salcedo-Sanz), prabal.barua@unisq.edu.au (P.D. Barua), rajendra.acharya@unisq.edu.au (U. Rajendra Acharya).

https://doi.org/10.1016/j.apenergy.2024.123920
Received 17 March 2024; Received in revised form 7 July 2024; Accepted 12 July 2024

such as weather conditions, availability of affordable power generation options like nuclear and hydroelectric, and unexpected issues in power generation and transmission can cause fluctuations in *EP* [4]. Therefore, accurate prediction methods to predict *EP* within the electric power industry is crucial.

Advanced technologies for *EP* prediction have been developed through numerous research efforts aimed at achieving highly accurate results. A significant amount of literature has been dedicated to *EP* prediction models, which can be classified into six categories: multi-agent [5], fundamental [6], reduced-form [7], statistical [8], computational intelligence (CI) models [9], and hybrid models [10]. Statistical, CI and hybrid models are commonly employed for *EP* prediction. To predict *EP* accurately, statistical methods such as Exponential Smoothing techniques, Auto Regression [11], Autoregressive Integrated Moving Average (ARIMA) [12], and Generalized Autoregressive Conditional Heteroskedastic (GARCH) [13] are commonly used, but they have limitations in predicting the nonlinear behaviour of *EP* [14]. Therefore, CI and hybrid models have gained popularity due to their superior performance in EP prediction. Hybrid Deep Learning models have been recognized as the most advanced CI approach in various fields and are now being applied in scientific research related to *EP* prediction.

Some popular CI methods utilized for predicting *EP* include Back Propagation Neural Network (BPNN) [15], Radial Basis Function (RBF) [16] Neural Network, Artificial Neural Network (ANN) [17], Bayesian Neural Network (BNN) [18], Support Vector Machine (SVM) [19], Random Forest (RF) [20], Least Squares Support Vector Machine (LSSVM) [21], Adaptive Neuro Fuzzy Inference System (ANFIS) [22], Wavelet Neural Network (WNN) [23], General Regression Neural Network (GRNN) [24], and Extreme Learning Machine (ELM) [25]. BPNN is prone to local optima, leading to limitations in accurately predicting *EP*. As an improved version of BPNN, ANN is widely used in *EP* prediction field. Keles et al. [26] developed a day-ahead *EP* prediction model based on ANN, which demonstrated good performance. To account for the impact of various external factors on *EP*, the input variables in the prediction model included hourly electricity price, hourly load value, hourly power generation of renewable energy, fuel price, available capacity divided by fuel type, holidays, and seasons. Ghayekhloo et al. [18] proposed a hybrid *EP* prediction model based on data clustering and Bayesian recursive neural network. The prediction model clusters input load and historical price data into a certain number of subsets and employs the Bayesian Recurrent Neural Networks (BRNN) model for *EP* prediction. The proposed clustering method has a Mean Square error value that is 13.7%, 14.9%, and 12.5% lower than that of K-means, Neural Gas, and self-organizing Map, respectively. Rafiei et al. [27] introduced a WNN and Generalized ELM (GELM) based *EP* prediction model. The WNN algorithm decomposes the electricity price sequence into sub-sequences with different frequencies using wavelet transform (WT), followed by the application of the GELM algorithm to achieve accurate forecasting in the Australian electricity market. However, with the entry of *EP* datasets into the big data field, the aforementioned shallow models face limitations in extracting deep nonlinear features from massive *EP* data due to their limited representation of complex functions with big data samples and computational units [28]. This has led to an increased interest in Deep Learning models for *EP* prediction, which have shown exceptional performance in language modelling, speech recognition, and natural language inference [29]. Common deep learning models used in *EP* prediction include Deep Neural Network (DNN) [30], Deep Belief Network (DBN) [31], Recurrent Neural Network (RNN) [32], Gated Recurrent Unit (GRU) [33], Long Short-term Memory (LSTM) [34], and Convolution Neural Network (CNN) [35].

The study of Lago et al. [14] proposed a DL models (DNN, LSTM, and GRU) for predicting hourly *EP*. Their findings demonstrate that the suggested model achieves a significantly higher predictive accuracy compared to other benchmark models, such as MLP, CNN, and RBF. Moreover, among the three DL models, the DNN outperforms the other two with a statistically significant difference in accuracy. However, all three models are necessary to obtain a highly performing forecaster, as the GRU and LSTM models still exhibit superior accuracy at certain specific hours. Chinnathambi et al. [36] introduced a DNN model for predicting daily *EP* in the Iberian electricity market (MIBEL), serving mainland regions of Spain and Portugal. The findings indicate that the DNN model performs exceptionally well in predicting *EP* compared to MLP.

The study of Jiang and Hu [37] recommended a LSTM model for predicting day-ahead electricity prices in both the Australian market at Victoria (VIC) region and the Singapore market. The findings demonstrate that the LSTM network surpasses three commonly used prediction model (ANN, ANFIS, and SARIMA) and delivers a considerable improvement of up to 47.3% in the average daily Mean Absolute Percentage Error (MAPE) for the VIC market. In a similar manner, Zhu et al. [38] utilized hourly price data from the New England and PJM day-ahead markets to develop an LSTM model. The performance of the LSTM model was also compared to other existing methods, including Decision Trees (DT) and Support Vector Machines (SVM). The predicted results reveal that the LSTM model generates a lower Root Mean Square Error (RMSE; USD/MWh) of approximately 36.05, in contrast to 37.51 and 49.15 for DT and SVM, respectively.

The study of Mujeeb et al. [39] introduced an LSTM model for the prediction of short-term load and price. The framework consists of data preprocessing, training an enhanced LSTM model, and predicting load and price patterns for 24 and 168 h. The practicality and feasibility of the proposed LSTM model were verified by evaluating its performance on established real market data from Independent System Operator, New England. Zahid et al. [40] developed Enhanced CNN for electricity load and price prediction. The model proposed in the study addressed the issue of over-fitting and effectively reduced execution and computation time. Ugurlu et al. [32] employs GRU, a variant of RNN, for *EP* prediction. The GRU model demonstrates superior accuracy in comparison to LSTM and various statistical time series prediction models.

In spite of several successful studies, there is a clear literature evidence that a single predictive model without a dedicated data preprocessing or pattern recognition scheme may struggle to accurately capture the intricate relationships in non-linear and non-stationary *EP* data series [41–43]. As a result, researchers have increasingly turned their attention to the development of hybrid predictive models. These models are sophisticated forecasting frameworks that integrate multiple algorithms. Typically, hybrid models consist of at least two of the following five modules [44]: (a) one or more prediction models with combined prediction (b) time-series decomposition algorithm, (c) feature selection algorithm, (d) data clustering algorithm, and (e) heuristic optimization algorithm to determine the optimal parameters of the model.

A popular hybrid method for predicting *EP* involves combining a standard Multilayer Perceptron (MLP) [45] with other methods such as Adaptive Neuro-Fuzzy Inference System (ANFIS) [46], Radial Basis Function (RBF) [47], and Autoregressive Moving Average (ARMA) or Autoregressive Integrated Moving Average (ARIMA) [48] or generalized regression neural network and gravitational search algorithm [49]. Additionally, there are other models that utilize a combination of Convolutional Neural Network (CNN) [50], Long Short-Term Memory (LSTM) [51], Linear Regression (LR) [50], Extreme Learning Machine (ELM) [52], Bayesian Neural Network (BNN) [53], Echo State Network (ESN) [54], Elman Neural Network [55], and Support Vector Machine (SVM) [56] to predict *EP*.

Recently, the integration of CNN with the other models is becoming highly popular and widely applied. The CNN layer plays a crucial role as an information feature extraction module, generating feature vectors of varying scales. These feature vectors serve as input for subsequent predictive models, enhancing their performance and accuracy.

Furthermore, as previously mentioned, *EP* series exhibit a blend of various characteristics, including short-term periodicities, long-term trends, non-constant mean and variance, and random spikes.

Improving prediction accuracy solely through the use of the hybrid model (combining two models) becomes challenging without considering the decomposition of these data characteristics within the *EP* series. The second category of hybrid models endeavours to address this issue by employing data decomposition pre-processing methods to alleviate the detrimental impact of the aforementioned characteristics. Through the decomposition process, a *EP* series is transformed into a set of constituent series, which display more desirable behaviour such as increased stability in variance and absence of outliers compared to the original price series. As a result, these constituent series can be predicted with higher accuracy.

There are several popular techniques available for decomposing time-series data. These techniques include Wavelet Transform (WT) [57], Discrete WT (DWT) [58], Empirical Mode Decomposition (EMD) [59], Ensemble EMD (EEMD), and variations of EMD such as Complete EEMD with Additive Noise (CEEMDAN) [60]. Maximal Overlap DWT (MoDWT) [61] is an advanced version of WT that improves its decomposition capabilities. A more recent version of EMD called improved CEEMDAN (ICEEMDAN) [62] has superior decomposition abilities compared to previous versions. Another technique, Variational Mode Decomposition (VMD) [63], has a proven track record of effectively decomposing time-series data for various applications.

It must be noted, however, time-series data can be influenced by both deterministic and stochastic effects, which can impact the accuracy of prediction models. Time-series decomposition allows for the separation of these components by breaking down the data into sub-components. By summing these sub-components, the original time-series can be reconstructed. MoDWT, ICEEMDAN, and VMD are capable of directly handling any type of time-series data, regardless of its stationarity, linearity, or stochastic nature. Consequently, these decomposition techniques have been widely used for feature extraction, noise reduction, and filtering in numerous studies [61,62].

Considering the complexity of EP prediction as a real-world, energy market problem, there is a growing need for accurate and efficient models. To address this, Ghimire et al. [2] developed a two-step deep learning framework with error compensation for half-hourly electricity price forecasting using Australian National Electricity Market data of five different regional market jurisdictions, namely Queensland, New South Wales, Victoria, South Australia, and Tasmania. That study demonstrated variational mode decomposition, coupled with an error compensation strategy as an effective means to predict half-hourly electricity prices. This study now focuses on developing both a decomposition and an ensemble deep hybrid model, denoted as **MoDWT-CRVFL**, for half hourly *EP* prediction, albeit using data for New South Wales, Australia. The proposed model combines a robust time series decomposition capability known as the MoDWT technique based on strong feature extraction capability of the CNN, and the powerful nonlinear fitting predictability of the Random Vector Functional Link (RVFL).

By integrating the VMD and CNN-RVFL algorithms, this study captures the merits of each standalone algorithm with an overall aim to improve the accuracy of *EP* prediction. This study therefore presents several significant contributions to knowledge and application in the area of electricity price forecasting as follows:

(a) The introduction of MoDWT to decompose the original *EP* time-series into multiple stationary components, taking into consideration the volatility and nonlinear characteristics of electricity prices more comprehensively than simplistic models.
(b) The utilization of antecedent *EP*, extracted using the Partial Autocorrelation Function (PACF) whereby input features are based on historical behaviour of electricity prices for predicting the future electricity price values.
(c) The proposal of a deep learning hybrid prediction algorithm called the CNN-RVFL. Compared to many other benchmark algorithms, this proposed deep learning hybrid algorithm demonstrates notable improvements in prediction accuracy and result stability, while also accelerating the training time and model convergence.
(d) The proposed MoDWT-CRVFL model, with its hyperparameters optimized using Bayesian Hyperparameter Optimization, is employed to model half-hourly *EP* scenarios. The effectiveness of the model is validated in four different test seasons: Summer, Winter, Autumn, and Spring.

To conduct these validations, a recent half-hourly *EP* dataset from New South Wales, Australia, collected between January 2014 and December 2022 is utilized. This dataset accounts for the uncertainties stemming from the COVID-19 pandemic and considers the impacts of ongoing political and energy-related crises within the Australian context.

This study is organized as follows: Section 2 shows the theoretical background of the techniques used to construct CRVFL, including MoDWT decomposition techniques, is described. Section 3 provides a description of the data used and outlines the steps taken to develop prediction models. This section also includes prediction metrics used to evaluate the performance of the constructed models. Section 4 focuses on presenting and discussing the results of the study, including a comparative analysis of prediction outcomes. Section 5 serves as the conclusion of the study, summarizing the main findings and conclusions. Appendix A shows the list of acronyms used in this study.

2. Theoretical background

This section provides a concise overview of the related methods, including the data preprocessing technique Maximum Overlap Discrete Wavelet Transform (MoDWT), Convolutional Neural Networks (CNN), Random Vector Functional Link Networks (RVFL), and the formulation process of the MoDWT-CRVFL model.

2.1. Maximal overlap discrete wavelet transform

In this study, the Maximal Overlap Discrete Wavelet Transform (referred to as MoDWT hereafter) has been utilized to decompose the electricity price time-series into different levels of scales and to extract the signal's variability at each scale [64]. The MoDWT, similar to the Discrete Wavelet Transform (DWT), is a type of linear filtering process. However, it is a non-orthogonal transform, which distinguishes it from the DWT. The MoDWT's fundamental concept involves using the values removed from the DWT through down sampling [65]. As a result, the MoDWT is highly redundant compared to the DWT, as it is defined for all sample sizes. Like the DWT, the MoDWT is useful for performing multiresolution analyses (MRAs), and the redundancy of the MoDWT allows for immediate comparison between the original time series and its decomposition at each level.

It is imperative to mention that the MoDWT coefficients of various scales are often uncorrelated, making it a useful transform for partitioning signal variability. Additionally, MoDWT-based algorithms provide more versatility compared to Empirical Mode Decomposition (EMD)-based methods as they can be tailored to suit specific applications and are based on a well-defined mathematical framework that is easily comprehensible. Unlike EMD, they do not require prior empirical estimation or significant modifications to the sifting procedure [66]. Moreover, the EMD-based method can result in mode mixing, which occurs when an intrinsic mode function contains components of different frequencies [67].

The decomposition process of MoDWT is depicted in Fig. 1. The MODWT-based decomposition involves mathematical equations

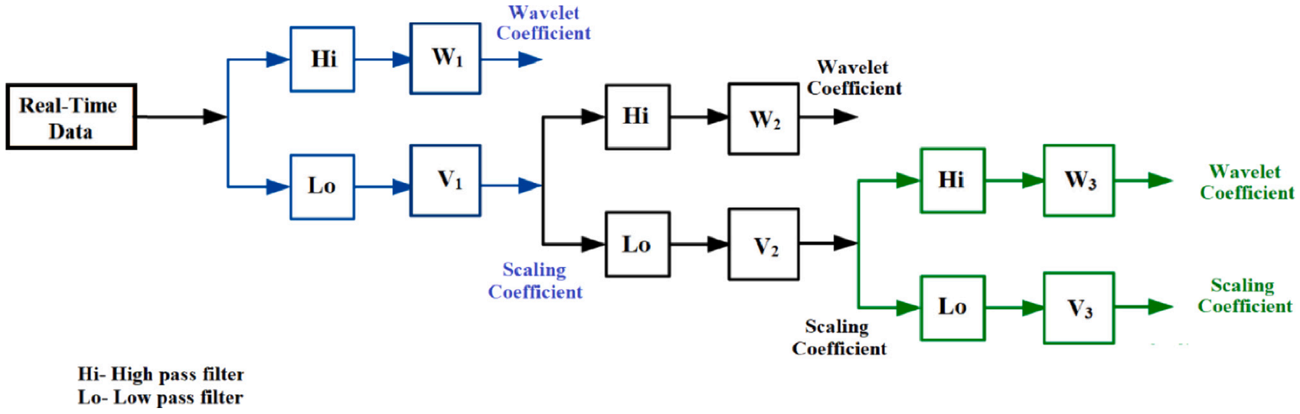


Fig. 1. The decomposition process of Maximal Overlap Discrete Wavelet Transform (MoDWT).

to compute wavelet (W) and scaling (V) coefficients for a discrete signal $Y = \{Y_t = 0, \dots, N\}$, which can be expressed as:

$$\text{Wavelet coefficient} \rightarrow \overline{W}_{jl} = \sum_{l=0}^{L_j-1} \tilde{h}_{jl} Y_{t-l\ (\mathrm{mod}\ N)} \tag{1}$$

$$\text{Scaling coefficient} \rightarrow \overline{V}_{jl} = \sum_{l=0}^{L_j-1} \tilde{g}_{jl} Y_{t-l\ (\mathrm{mod}\ N)} \tag{2}$$

where $L_j = (2^j - 1)(L - 1) + 1$, $\tilde{h}_{jl}$ represents the wavelet filter (high-pass) and $\tilde{g}_{jl}$ represents the scaling filter (low-pass) for the jth level of the MODWT. Additionally, j indicates the level of decomposition, and l represents the length of the signal. The wavelet filter (high-pass) and scaling filter (low-pass) are calculated using Eqs. (3) and (4):

$$\tilde{h}_{jl} = \frac{h_{jl}}{2^{\frac{j}{2}}} \tag{3}$$

$$\tilde{g}_{jl} = \frac{g_{jl}}{2^{\frac{j}{2}}} \tag{4}$$

where h_{jl} and g_{jl} are derived from the jth level of DWT's high-and low-pass filters. Thus, the MoDWT coefficients at level j are defined as the convolutions of the original time series $\{Y\}$, and it has the same length as that of $\{Y\}$. Meanwhile, Eqs. (1) and (2) can be expressed in matrix form as below:

$$\widetilde{W}_j = \overline{\overline{w}}_j Y \tag{5}$$

$$\widetilde{V}_j = \overline{\overline{v}}_j Y \tag{6}$$

where the $N \times N$ matrices $\overline{\overline{w}}_j$ and $\overline{\overline{v}}_j$ comprises of the wavelet coefficients $\tilde{h}_{jl}$ and scaling coefficients $\tilde{g}_{jl}$, respectively. Thus, the original signal $\{Y\}$ can be represented by Eq. (7):

$$Y = \sum_{j=1}^{L} \overline{\overline{w}}_j \widetilde{W}_j + \overline{\overline{v}}_j \widetilde{V}_L = \sum_{j=1}^{L} D_l + S_L \tag{7}$$

where the details coefficient $D_l = \overline{\overline{w}}_j \widetilde{W}_j$ denotes the irregular fluctuations of the series $\{Y\}$ at scale $l \ \forall \ (l = 1, 2, \dots, L)$ and the smooth $S_L = \overline{\overline{v}}_j \widetilde{V}_L$ indicates the overall trend of the original signal at scale L [68]. In addition, works [61,69,70] provide more detailed information about the MoDWT and its applications in various fields. For a comprehensive mathematical explanation, readers can refer to the Ref. [71].

2.2. Convolution neural network

This study has adopted the Convolution Neural Network (CNN), a well-known Artificial Neural Network (ANN) algorithm for Deep Learning due to its popularity and its ability to perform nonlinear feature extraction [72]. Various CNN architectures are available in the literature, which are created by modifying the number of layers including convolutional, pooling, and fully connected (dense) layers. The convolutional layer uses filters or neurons composed of kernels to generate feature maps from the input data [73]. Each kernel creates a unique weight matrix that is used to generate a feature map, and the weight values and bias term are adjusted during the training phase. The literature typically discusses two types of CNN, two-dimensional (2D-CNN) and one-dimensional (1D-CNN), which differ in the dimension of the filter. 2D-CNN is commonly used in image and text analyses, while 1D-CNN is frequently used with numerical data like energy production and meteorological variables. Despite this difference, both types of CNN have the same main layers, with variations in the filter's dimension [74].

Fig. 2 shows an example of a 1D-CNN with all four primary layers.

The equation representing the mathematical formula for the convolutional layer is displayed in Eq. (8):

$$y^l_{i,j,k} = \sigma\left(\left(w^l_k\right)^T x^l_{ij} + b^l_k\right) \tag{8}$$

where the weight and bias of kth convolutional kernel in the lth layer are represented as w^l_k and b^l_k, respectively. x^l_{ij} is the input patch in the lth layer, concentrated at the location (i, j). σ represents the activation function, e.g. ReLU, Sigmoid, Tanh, etc., to achieve nonlinearity [75]. The primary objective of the pooling layer is to reduce the resolution of the feature map, and it is often located between two convolutional layers [76]. The mathematical depiction of the pooling layer is presented in Eq. (9) [77]:

$$P^l_{i,j,k} = \mathrm{Pool}\left(y^l_{m,n,k}\right) \tag{9}$$

In Eq. (9), the notation $(m, n) \in R_{ij}$ denotes the area around the location (i, j). The specific type of pooling operation used in the layer, such as average pooling or max-pooling, is indicated by the $\mathrm{Pool}(\cdot)$ function. Typically, the pooling layer enhances network accuracy and reduces training time by decreasing the number of parameters in the network.

The fully connected layer, which is also known as the dense layer, is designed to perform high-level reasoning by transferring the learned features in the network to a unified space [78]. This is accomplished by linking every neuron in the current layer with each neuron from the previous layers. In most CNN models, one or more dense layers are found after the convolution and pooling layers. The final dense layer generates the network output. Additionally, as depicted in Fig. 2, a flattening layer exists between the CNN layer (which includes the Convolution layer and Pooling layer) and the Dense layer. The Flattening layer is employed to convert the multi-dimensional feature maps obtained from the preceding pooling layer into a one-dimensional array that satisfies the data processing needs of the subsequent dense layer [79].

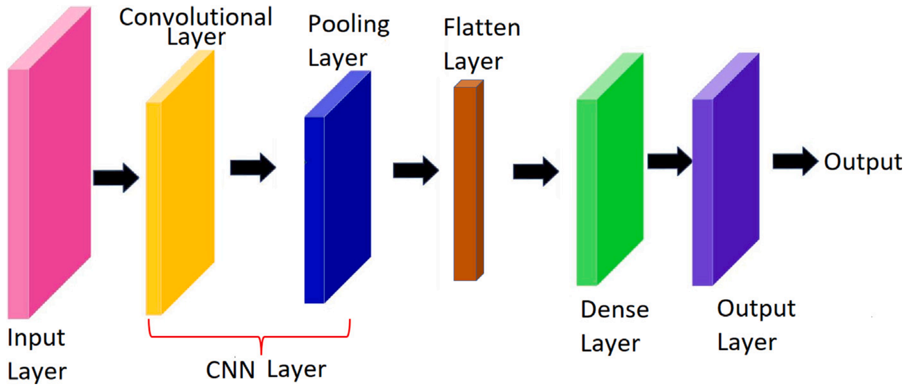


Fig. 2. The architecture of 1D-CNN model. This model consists of one convolutional layer,one pooling layer, one flatten layer, one dense layer and one output layer.

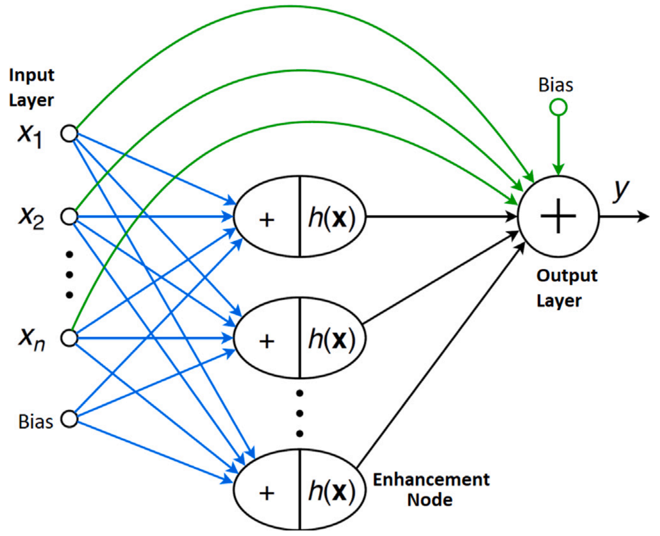


Fig. 3. The architecture of RVFL Network. Note:- random links in blue, direct links and output node bias in green.

2.3. Random vector functional link networks

In addition to the CNN algorithm, this study has utilized Random Vector Functional Link (RVFL) network, introduced by Pao and Takefuji [80]. The RVF can be viewed as a particular instance of Single Hidden Layer Feedforward Neural Networks (SLFNN). In the RVFL network, the input weights and hidden biases are randomly determined (no tuning required), and the other network parameters are acquired through a pseudoinverse [81]. In addition to this, the RVFL network includes direct links that connect the input layer to the output layer [82]. These features enable the RVFL network to achieve high efficiency in comparison to traditional iterative learning techniques and non-iterative networks, such as ELM [83], that lack direct input–output links. Therefore, the non-iterative RVFL network is integrated with CNN and used as a prediction tool to improve the performance of existing decomposition-ensemble models, in terms of both accuracy and speed.

The RVFL structure is illustrated in Fig. 3.

The RVFL network comprises three types of nodes: input nodes, output nodes, and enhancement nodes. Enhancement nodes are similar to the nodes in the hidden layer of a SLFNN and are used to map the input data from the input layer to the hidden layer. A brief description of each layer is presented below.

- Input Layer: The primary role of the input layer is to receive a training dataset $\{(x_i, y_i)\}$ consisting of n samples, where i ranges from 1 to n, and x and y are both elements of the real numbers, i.e., $x \in \mathbb{R}^n, y \in \mathbb{R}$.
- Hidden Layer: The hidden layer in a neural network calculates the activation function value ($h(\cdot)$) for each of its nodes, and typically uses the sigmoid function to compute this value, which is expressed in Eq. (10).

$$h(x, w, b) = \frac{1}{1 + \exp\{-w^T x + b\}} \tag{10}$$

where w and b are weight and biases from the input layer to the hidden layer, respectively. Using Eq. (11), the output can then be calculated by creating a kernel mapping matrix (H) for the hidden layer.

$$H = \begin{bmatrix} h_1(x_1) & \cdots & h_k(x_1) \\ \vdots & \ddots & \vdots \\ h_1(x_n) & \cdots & h_k(x_n) \end{bmatrix} \tag{11}$$

where k is the number of hidden layer nodes
- Output Layer: The primary objective of training RVFL networks is to determine the optimal weights ($\mathbf{W_o}$) from the hidden layer to the output layer. These weights can be computed using the least-square approach and solved through the following process:

$$\mathbf{W_o} = (H^T H)^{-1} H^T Y \tag{12}$$

where Y is the training target.

As previously stated, random values within the range $[-K, K]$ are assigned to all hidden layer weights ($\mathbf{W_h}$), and the value of K must be optimized. Finally, during the testing phase, the trained optimal weights ($\mathbf{W_o}$) and ($\mathbf{W_h}$) can be utilized to predict values for the testing data.

$$\hat{Y}_s = \mathbf{W}_o \cdot \sigma(\mathbf{W}_h \cdot \mathbf{X}_s) \tag{13}$$

where $\hat{X}_s$ and $\hat{Y}_s$ are the testing data and corresponding predicted values, respectively. σ is the activation function.

2.4. Convolutional neural network integrated with random vector functional link

The proposed MoDWT-CRVFL model comprises of the CRVFL network, a hybrid model that utilizes the benefits of both CNN and RVFL. It leverages the high-dimensional features extracted by the 1-D CNN, which is well-suited for temporal feature representation, and replaces the fully connected layer with RVFL, which uses a fast non-iterative algorithm. Unlike traditional CNNs, which are good at processing spatial autocorrelation data but struggle with complex and long-term time dependence [84], the CRVFL network can better handle nonlinear systems and improve prediction performance.

Fig. 4 shows the CRVFL architecture with an input layer, a convolution layer, a pooling layer, a flattening layer, and a RVFL network. The

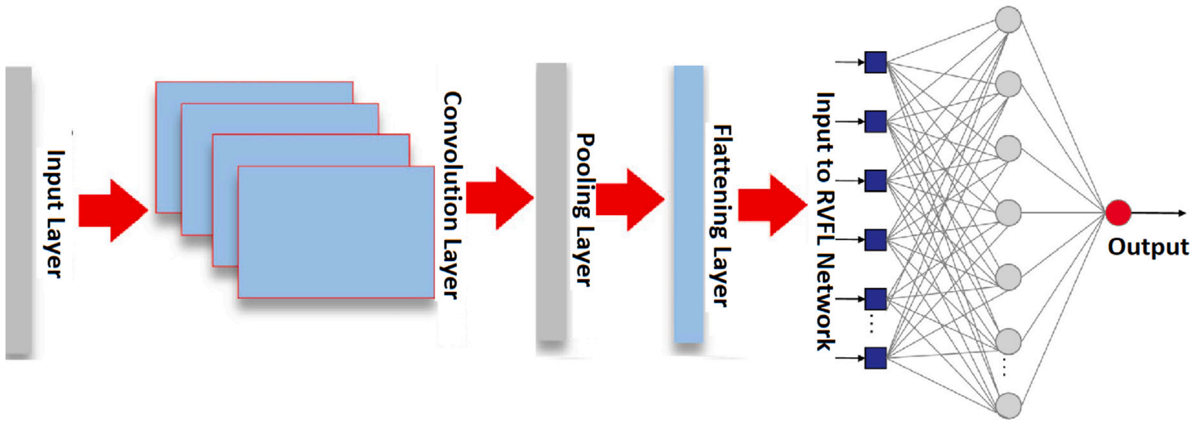


Fig. 4. The architecture of CNN integrated with RVFL Network.

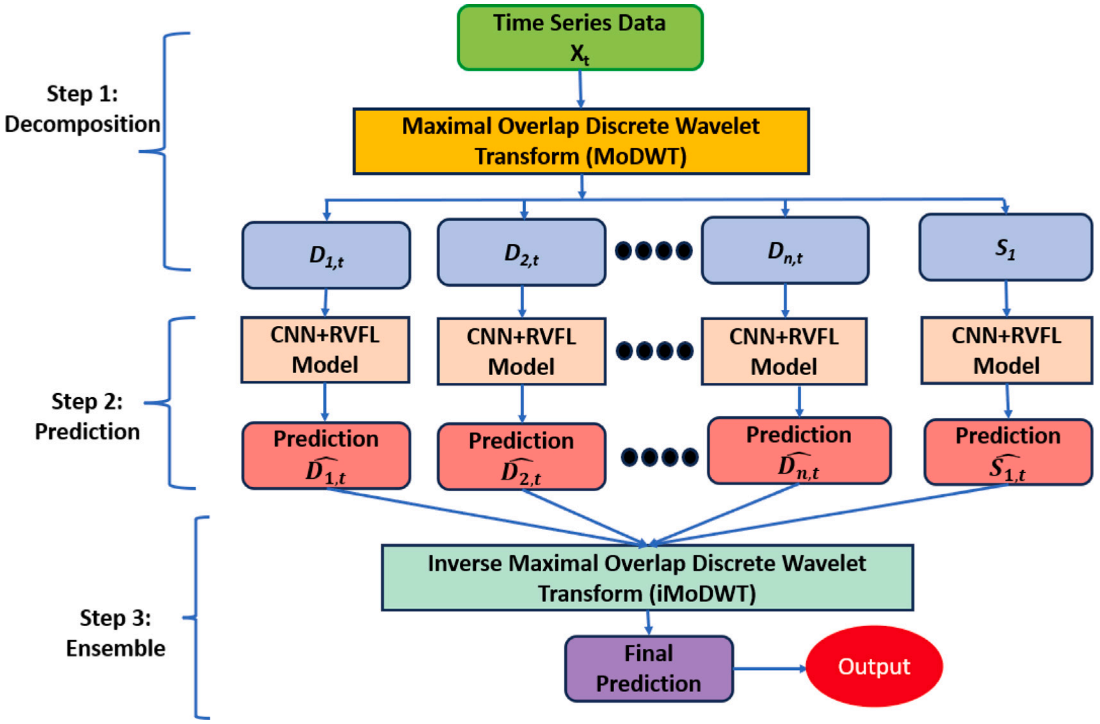


Fig. 5. Model framework of the proposed decomposition-based deep hybrid learning paradigm.

input layer receives the time-series data and feeds it to the CNN layer for high-level feature extraction. Next, the output from the pooling layer is flattened and sent to the RVFL layer, which produces the final prediction.

2.5. Decomposition and ensemble framework

To develop the proposed objective (i.e., MoDWT-CRVFL) model, the decomposition and ensemble framework is illustrated in Fig. 5. This comprises of three major steps. Firstly, the MoDWT is used to extract multiple components of the *EP* time-series in the decomposition step.

Next, these components are predicted using the CRVFL in the prediction step. Finally, the prediction results are inverse transform to obtain the final prediction. The fundamental idea behind this framework is to simplify the problem by breaking down the complex *EP* time-series into several sub-problems, solving them, and combining the results to achieve a more accurate prediction.

- Decomposition: The original time-series $x_t (t = 1, 2, 3, \dots, T)$ was decomposed into several (n) detail coefficients $D_{j,t} (j = 1, 2, 3, \dots, n)$ and 1 level of scaling coefficient (S_1) using MoDWT.
- Prediction: The CRVFL network is utilized as a prediction tool to model each of the scaled and detail coefficient. As a result, individual prediction result $\hat{D}_{j,t}$ and $\hat{S}_1$ can be obtained for each of the corresponding components via inverse MoDWT.
- Ensemble: Inverse MoDWT is applied to individual results to produce the final results.

2.6. Benchmark models

The performance of the proposed MoDWT-CRVFL model was evaluated using competing benchmark models: Multilayer Perceptron (MLP), Random Forest (RF), Extreme Gradient Boost (XGB), deep Neural Network (DNN), Long Short-Term Memory (LSTM) Network and the Bi-Directional LSTM (BLSTM). An MLP model is a powerful data-driven modelling tool with three layers [76]: an input layer, a hidden layer, and an output layer. This model operates as a fully black-box system, taking inputs and producing the desired output for the provided dataset. In contrast to MLP, DNN exhibits a deeper and more complex network structure with additional hidden layers and weight connections [85,86].

The RF and XGB models belong to the category of ensemble learning algorithms. Ensemble learning is a methodology that seeks to improve the accuracy of final prediction values by constructing multiple models and aggregating their respective predictions [87]. The mathematical interpretation of RF and XGB can be found in Ref. [88] and Ref. [89], respectively. LSTM, developed by Hochreiter and Schmidhuber [90], is an advanced variation of the RNN network that overcomes its limitations by incorporating memory cells within the hidden layer. These memory cells store the temporal state of the network and are controlled by three gates: the input gate, output gate, and forget gate [62,91].

The Bidirectional LSTM (Bi-LSTM) model has also been employed as a baseline model for *EP* prediction. While LSTM can only learn from past information, Bi-LSTM model has the capability to learn from both past and future information simultaneously. This is achieved by incorporating two hidden layers with opposing directions that connect to the same output [62].

3. Material and method

This section details the characteristics and preprocessing of the electricity price data, the systematic development of the prediction model, the implementation of benchmark models for comparative analysis, and the specific metrics used to evaluate model performance. These components collectively provide a thorough understanding of the study's methodology and evaluation criteria

3.1. Electricity price dataset

The evaluation of the proposed MoDWT-CRVFL model was undertaken using electricity price (*EP*) data series obtained from the Australian electricity market (AEM: https://www.aemo.com.au/) for the New South Wales (NSW) region. In general, the AEM provides 48 daily observation values with a time interval of 30 min between each point. A total of 122,735 observation values spanning from January 1, 2016, to December 31, 2022 were analysed to evaluate the effectiveness of the proposed model.

Fig. 6 presents box-plots that demonstrate the significant influence of various factors on *EP*, including the day of the week, month of the year, time of the day, and year. It is noticeable that the price patterns align with the expected human behaviour and industrial and commercial practices. A clear trend emerges when examining the relationship between *EP* and the hours of the day. During the early hours, *EP* tend to be lower, reflecting reduced demand and typical human behaviour.

Notably, *EP* are generally lower between 1:00 and 5:00, with the lowest point reached at 4:00. Conversely, higher *EP* are observed in the morning at 8:00 and in the late afternoon between 17:00 and 19:00. The box-plots for these peak hours in the late afternoon exhibit longer whiskers, indicating a higher variability in *EP* during that time period.

The analysis of box-plots reveals that *EP* are lower on weekends (Saturday and Sunday) compared to business days. Moreover, there is a seasonal trend in *EP*, with higher *EP* observed during the winter months (June, July, and August) in contrast to the summer months (December, January, and February).

Considering the seasonal variation characteristics in the *EP* time series from the AEM, it is important to assess the impact of different seasons on the accuracy and stability of the proposed MoDWT-CRVFL model. To address this, the *EP* dataset is divided into five parts, see Table 1 and Fig. 7.

Table 1 provides statistical information about various datasets including the mean (Mean), maximum (Max), and minimum (Min) values, the standard deviation (Std), skewness (Skew), and kurtosis (Kurt), enabling a comprehensive analysis of the data. All five dataset (*DS1*, *DS2*, *DS3*, *DS4* and *DS5*) exhibit kurtosis values higher than 3, indicating that the price distribution of electricity displays fat tails, which suggests a higher likelihood of extreme values. The *EP* series demonstrates a skewness greater than 1, indicating a significant skew in the distribution. Moreover, the rejection of the null hypothesis of normal distributions, as indicated by the Jarque–Bera statistic(JB_{stat} and $JB_{P_{val}}$) [92], provides further evidence that the price distribution deviates from a normal distribution pattern.

An Augmented Dickey–Fuller test (ADF Test) or a unit root test was conducted on the *EP* time series in order to determine its stationarity [93]. It appears that all five datasets reject the null hypothesis that the time series have a single root, indicating that they satisfy the condition of stationarity, see Table 2. As a further way to examine the linearity or nonlinearity of data, this study has applied the BDS [94] independence test. Table 3 shows that for all five datasets, the BDS independence test decisively rejected the null hypothesis that time series are linearly dependent. In this case, the BDS independence test indicates that *EP* shows highly nonlinear and chaotic patterns.

3.2. Development of the objective model

Step 1: Data Normalization When adapting Deep Learning-based predictive models, it is crucial to normalize the data, which is sometimes referred to as scaling. The primary objective of this process is to avoid any bias towards input variables that have larger numerical values compared to those with smaller values. Additionally, since kernel-based methods heavily rely on the inner multiplication of input vectors, computational challenges can arise when dealing with large-value input variables. Therefore, normalizing input vectors becomes essential to address these numerical complexities during the computation processes. In this study thee normalization of the half-hourly *EP* time-series are done using Eq. (14):

$$Z_t = \frac{Z_{i,t} - Z_{i,t}^{avg}}{Z_{i,t}^{std}} \tag{14}$$

where Z_t, $Z_{i,t}$, $Z_{i,t}^{avg}$ and $Z_{i,t}^{std}$ is normalized value, input data of time t for ith quantity, average value for ith quantity and standard deviation for ith quantity respectively.

Step 2: MoDWT based Decomposition Analysis of NSW *EP* time-series

To decompose the half-hourly *EP*, the MoDWT algorithm is utilized. The selection of an appropriate mother wavelet is a crucial factor in this decomposition process. While the MoDWT is less dependent on the specific mother wavelet filter compared to the DWT, it is still important to consider the diverse properties of different mother wavelet filters as they can influence the results of the wavelet analysis.

In this study, the energy to Shannon entropy ratio R_{ES} and energy to Renyi entropy ratio R_{ER} is used as the criterion for selecting the mother wavelet. The mother wavelet with the highest R_{ES} and R_{ER} is considered the most suitable for the MoDWT decomposition. For further details on the mathematical derivation and features of the R_{ES} and R_{ER}, see [95–99]. The EP time series was subjected to wavelet transform using various filters from commonly used wavelet families, including Daubechies ($db2$, $db4$, $db6$, $db8$, $db10$, $db12$, $db14$, $db16$, $db18$ and $db20$), Symlet ($sym4$, $sym5$, $sym6$, $sym8$, $sym9$ and $sym10$), Fejer-Korovkin ($fk6$, $fk8$, $fk14$, $fk18$ and $fk22$), discrete approximation of Meyer wavelet ($demy$) and Coiflet ($coif2$, $coif3$, $coif4$, $coif5$). In total, 26 filters were employed, all of which are commonly utilized in wavelet-based prediction.

Table 4 presents the R_{ES} and R_{ER} obtained after decomposing the *EP* time series using various wavelet bases. Prior to wavelet decomposition, the data is normalized utilizing Eq. (14). According

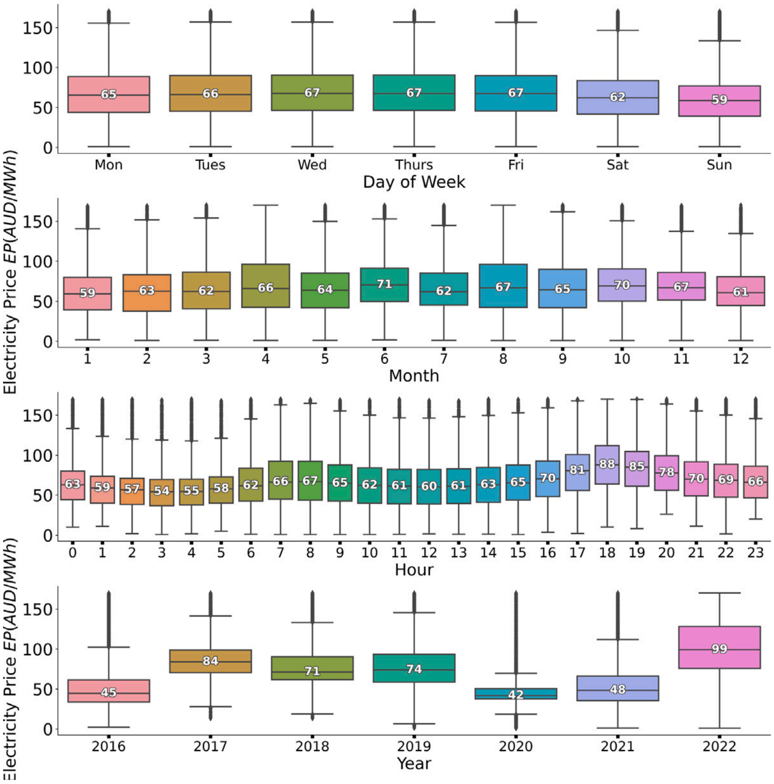


Fig. 6. Box plot of the electricity price (half-hourly) *vs.* day of the week, month of the year, Hour of the day and yearly, including median value (middle line), maximum (top whisker), and minimum (bottom whisker) values are shown for NSW electricity price data where the proposed MoDWT-CRVFL model was developed and evaluated. The value out of top and bottom whisker represents the outliers.

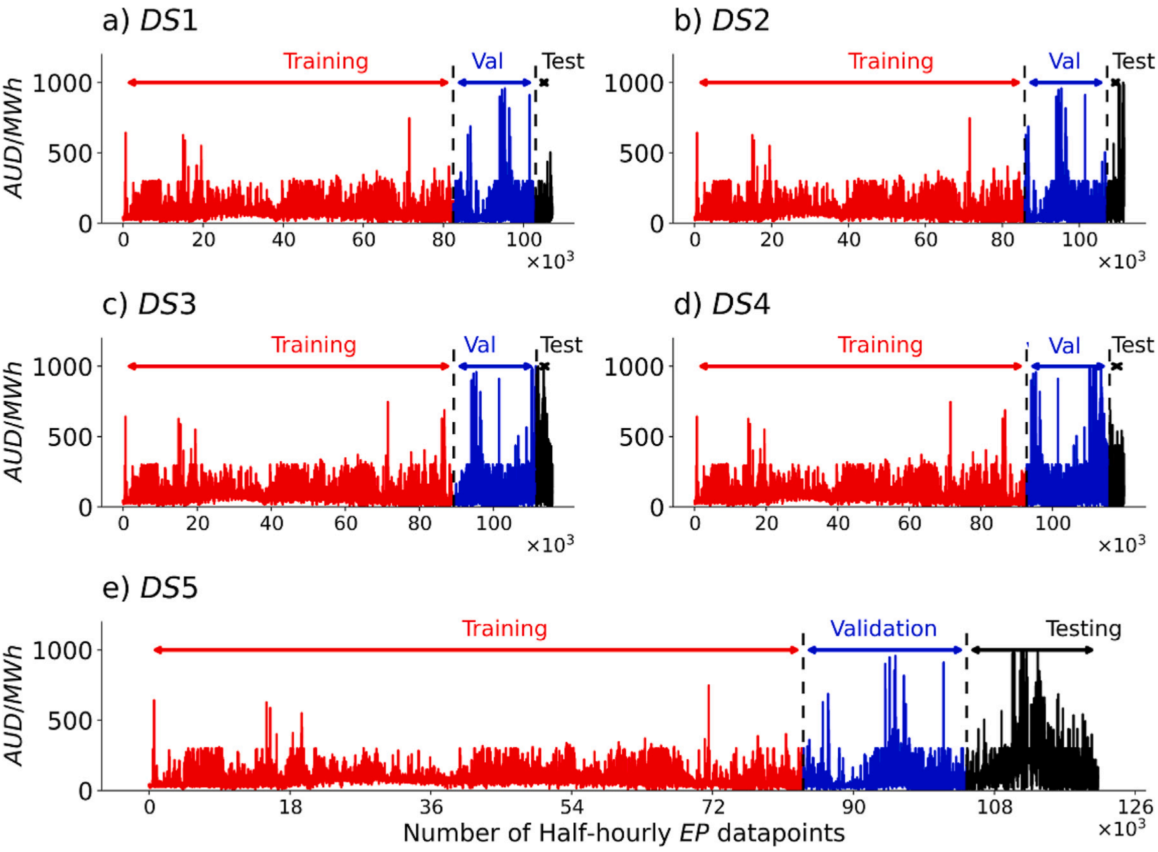


Fig. 7. Half-hourly electricity price data (2016 to 2022, see Table 1) where the proposed MoDWT-CRVFL model was developed. The objective (and benchmark) models were tested on datasets $DS1$ (summer), $DS2$ (autumn), $DS3$ (winter), $DS4$ (spring) and $DS5$ (year 2022) using the training, validation, and testing sets.

Table 1

Five test case scenarios (*i.e.*, Summer, Autumn, Winter, Spring and Yearly 2022) where the proposed MoDWT-CRVFL model for electricity prediction was comprehensively evaluated.

	Summer Season Scenario 1 = $DS1$	Autumn Season Scenario 2 = $DS2$	Winter Season Scenario 3 = $DS3$	Spring Season Scenario 4 = $DS4$	Year 2022 Scenario 5 = $DS5$
Training Dates	1-Jan-2016 to 30-Nov-2021	1-Jan-2016 to 28-Feb-2022	1-Jan-2016 to 31-May-2022	1-Jan- 2016 to 31-Aug-2021	1-Jan-2016 to 31-Dec-2021
Testing Dates	1-Dec-2021 to 28-Feb-2022	1-March-2022 to 31-May-2022	1-Jun-2022 to 31-Aug-2022	1-Sep-2021 to 31-Nov-2021	1-Jan-2022 to 31-Dec-2022
Training Samples	N = 82,442	85,805	89,319	92,759	83,577
Validation Samples	N = 20,610	21,451	22,330	23,190	20,894
Testing Samples	N = 4,204	4,393	4,300	4,043	16,785
Mean	68.73	84.92	108.30	81.88	63.23
Max	745.64	1000.00	997.70	910.63	958.14
Min	1.00	1.01	1.00	1.00	1.12
Std.Dev.	40.17	67.79	105.49	50.89	51.37
Skew	3.15	3.05	2.73	2.35	4.50
Kurt	20.35	15.95	9.01	10.18	40.02
JB_{stat}	565 838.29	373 924.59	141 720.92	156 397.41	1 186 386.99
JB_{Pval}	0	0	0	0	0

Mean, *Max*, *Min* and *Std. Dev* refers to the Mean, Maximum, Minimum and Standard Deviation of *EP* in *AUD/MWh* respectively. *Skew* refers to the Skewness, *Kurt* refers to the Kurtosis, whereas JB_{stat} = Jarque–Bera test [92] for normality and JB_{Pval} is the *p*-value at 95% confidence interval for the five datasets, N = data size, Validation data = 20% of training data.

Table 2

Augmented Dickey–Fuller test applied to check the presence of unit root to determine whether the model design data are non-stationary.

Statistical test parameter	*DS1*	*DS2*	*DS3*	*DS4*	*DS5*
ADF_{Stat}	−7.29	−2.80	−4.31	−6.22	−6.53
ADF_{Pval}	0.00	0.03	0.00	0.00	0.00

ADF_{Stat} = ADF statistic, ADF_{Pval} = *p*-value at the 95% confidence level.

to the criterion of selecting the wavelet with the maximum R_{ES} and R_{ER} values, the $db2$ wavelet exhibits the highest values among the options considered. Hence, the $db2$ wavelet is determined to be the most suitable choice for processing the half-hourly *EP* time series. Furthermore, the number of decomposition levels was determined using the equation $L = \log_2 N$, where N represents the length of the time series data.

In this study, all datasets (i.e., $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$) had a length (N) exceeding 90,000, leading to an approximate value of $L \approx 16$. Consequently, the half-hourly *EP* series was decomposed into 16 levels of the detail coefficients ($D_1, D_2, \ldots, D_{16}$) and 1 level of scaling coefficient (S_{16}) for each dataset.

Fig. 8 presents a graphical illustration of MoDWT decomposition applied to the $DS1$ training data. It is worth mentioning that the training and testing data (Fig. 7 and Table 1) are decomposed into distinct details and scaling coefficients to prevent future data from influencing the input series or introducing prediction biases [62].

Step 3: Significant time-lagged inputs extraction for model development: Following the MoDWT decomposition, the obtained detail coefficients (D_1, D_2, …, D_{16}) and scaling coefficient (S_{16}) are subjected to a stationarity assessment using the Augmented Dickey–Fuller (ADF) test.

The physical interpretation of the ADF test states that if a particular sub-series is found to be non-stationary, the first difference of that sub-series is examined for stationarity using the ADF test. This iterative process continues until either the sub-series achieves stationarity or the maximum number of iterations is reached.

In order to ascertain the relationship between a decomposed *EP* series and its past values, the Partial Auto Correlation Function (PACF) is employed. The PACF reveals the lag parameter, which indicates the dependence of the series on its previous values and aids in determining the inputs for the models.

Fig. 9 exhibits the PACF plot for $DS1$ acquired through MoDWT and Table 5 outlines the input parameters corresponding to each decomposed series within the datasets $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$, where D_1, D_2, …, D_{16} are detail coefficient and S_{16} is the scaling coefficient from MoDWT decomposition.

Step 4: MoDWT-CRVFL model Architecture The CRVFL model's architecture, shown in Fig. 10 and Table 6, represents a novel approach where the CNN and the RVFL methods are integrated to create a hybrid model, CRVFL, which is tailored for half-hourly *EP* prediction. This study harnesses the CNN model to extract spatial characteristics from multivariate time series data, comprising an input layer for variable input, a single output layer responsible for feature extraction for RVFL, and four hidden layers, each followed by pooling layers.

The CNN model processes lagged and normalized values from the decomposed *EP* series as input. Given that the input data is presented as a multivariate time series, specifically as a lagged matrix of MoDWT decomposed *EP* values, denoted as (X_{t-1}), (X_{t-2}), …, (X_{t-n}), this study represents the data as 3-dimensional tensors with dimensions (N, Q, M). In this notation, N signifies the number of samples, Q represents the maximum number of time steps across all variables, and M denotes the variables processed per time step. The value of N is determined by the number of features present in each dataset. For instance, when considering D_1 and D_2 from the $DS1$ dataset, the input layer receives 2 samples, namely (X_{t-1}) and (X_{t-2}).

Following the input layer, the first convolutional layer ($CL1$) is configured with $M1$ units (filters) and employs a kernel size of 1, utilizing the $ReLU$ activation function. Subsequently, the model integrates a second convolutional layer ($CL2$), a third convolutional layer ($CL3$), and a fourth convolutional layer ($CL4$), with each layer containing $M2$, $M3$, and $M4$ units, respectively. After each of these convolutional layers ($CL1$, $CL2$, and $CL3$), a pooling layer ($PL1$ and $PL2$)is introduced to effectively reduce the feature space. This process enhances the model's learning efficiency by producing more abstract training features.

Following the final pooling layer ($PL3$), a fourth convolutional layer is applied, followed by a flattening layer to transform the features

Table 3
The Brock, Dechert, Scheinkman and LeBaron (BDS) test of independence conducted to check whether the data series are linear or nonlinear.

Input at the Lag (m)	*DS1*	*DS2*	*DS3*	*DS4*	*DS5*
1	11.702 [0.0]	12.776 [0.0]	17.392 [0.0]	14.34 [0.0]	6.33 [0.0]
2	13.197 [0.0]	14.133 [0.0]	19.655 [0.0]	16.338 [0.0]	7.008 [0.0]
3	14.638 [0.0]	15.505 [0.0]	22.105 [0.0]	18.395 [0.0]	7.608 [0.0]
4	16.151 [0.0]	16.966 [0.0]	24.865 [0.0]	20.635 [0.0]	8.196 [0.0]
5	17.794 [0.0]	18.552 [0.0]	27.997 [0.0]	23.12 [0.0]	8.791 [0.0]
6	19.586 [0.0]	20.295 [0.0]	31.583 [0.0]	25.925 [0.0]	9.41 [0.0]
7	21.559 [0.0]	22.225 [0.0]	35.705 [0.0]	29.121 [0.0]	10.061 [0.0]
8	23.74 [0.0]	24.365 [0.0]	40.466 [0.0]	32.769 [0.0]	10.747 [0.0]
9	26.168 [0.0]	26.749 [0.0]	45.979 [0.0]	36.947 [0.0]	11.479 [0.0]

Note: m = number of lags with the critical value being 1.96 at the 95% confidence level. The null hypothesis is that the time series is linearly dependent. The value in [ ] indicates the corresponding p-value.

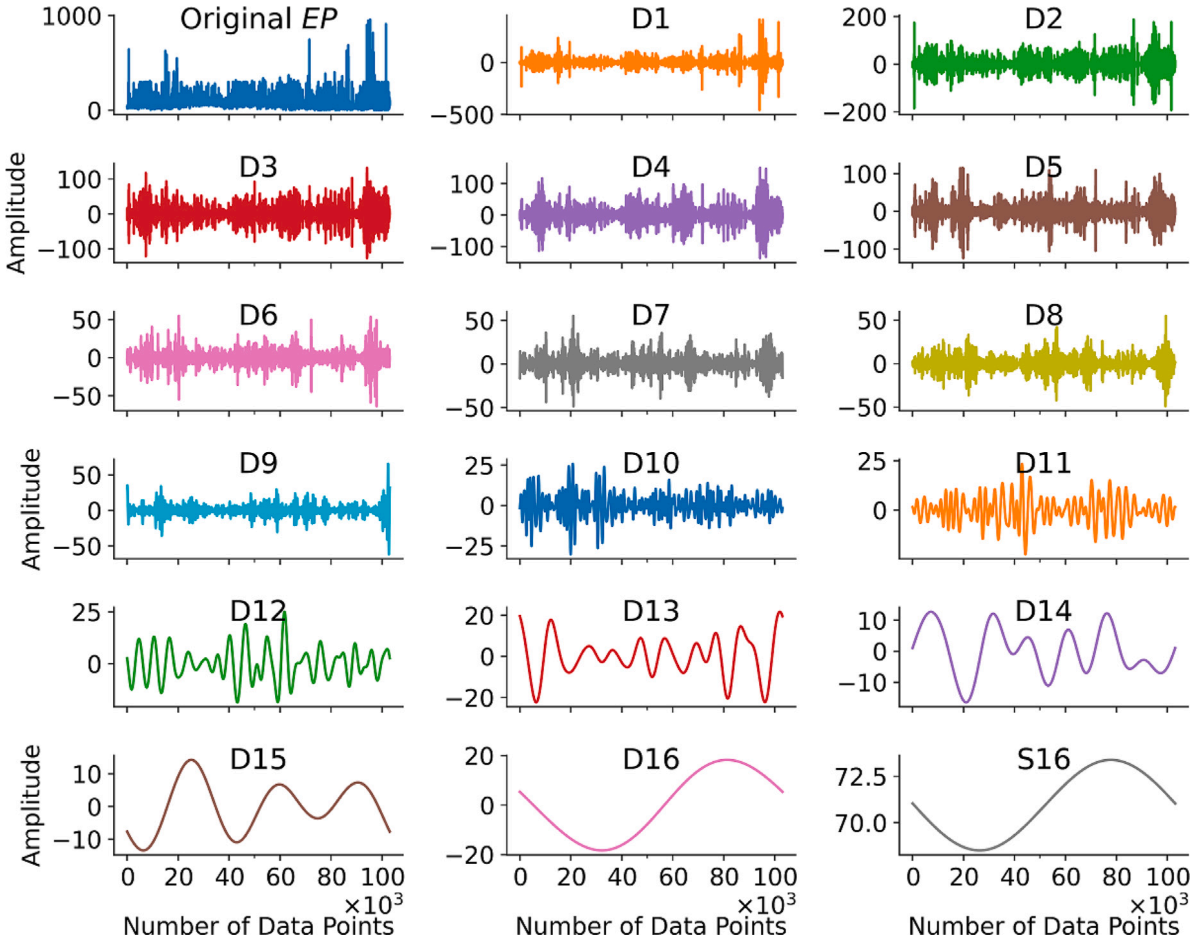


Fig. 8. The MoDWT-based decomposition of the *EP* dataset prior to developing the MoDWT-CRVFL model. The first curve represents the original time series; the last represents the smooth series (S_{16}), i.e., the scaling coefficient of the time series after computing the MoDWT algorithm. The other sixteen curves (D_1, D_2, …, D_{16}) represent the different level coefficient calculated by the MoDWT algorithm. The filter used for this decomposition is the $db2$ filter.

into a one-dimensional vector. This vector is subsequently fed into the RVFL model.

The RVFL model is instrumental in modelling the irregular time information using the spatial features it receives. It consists of three layers: an Input layer, an enhancement layer (hidden layer), and an output layer. Ultimately, the CRVFL model has the capability to generate predictions for *EP* in a fully connected hierarchy. The proposed CRVFL model predicts half-hourly-based *EP* data, which has been decomposed into D_1, D_2, …, D_{16} and S_{16} using MoDWT separately.

Step 5: Hyperparameter tuning and prediction using MoDWT-CRVFL model As previously stated, the proposed model employed lagged and normalized values from the MoDWT decomposed *EP* series as inputs. It is crucial to emphasize that the training data was utilized for model training, while a distinct 20% subset was set aside for validation, facilitating model refinement and performance evaluation, thereby reducing the risk of over-fitting (see Fig. 7 and Table 1).

The CRVFL model was utilized to predict both detail (D_1, D_2, …, D_{16}) and scaling coefficient (S_{16}). Moreover, hyperparameters constitute a critical set of parameters that define the architecture

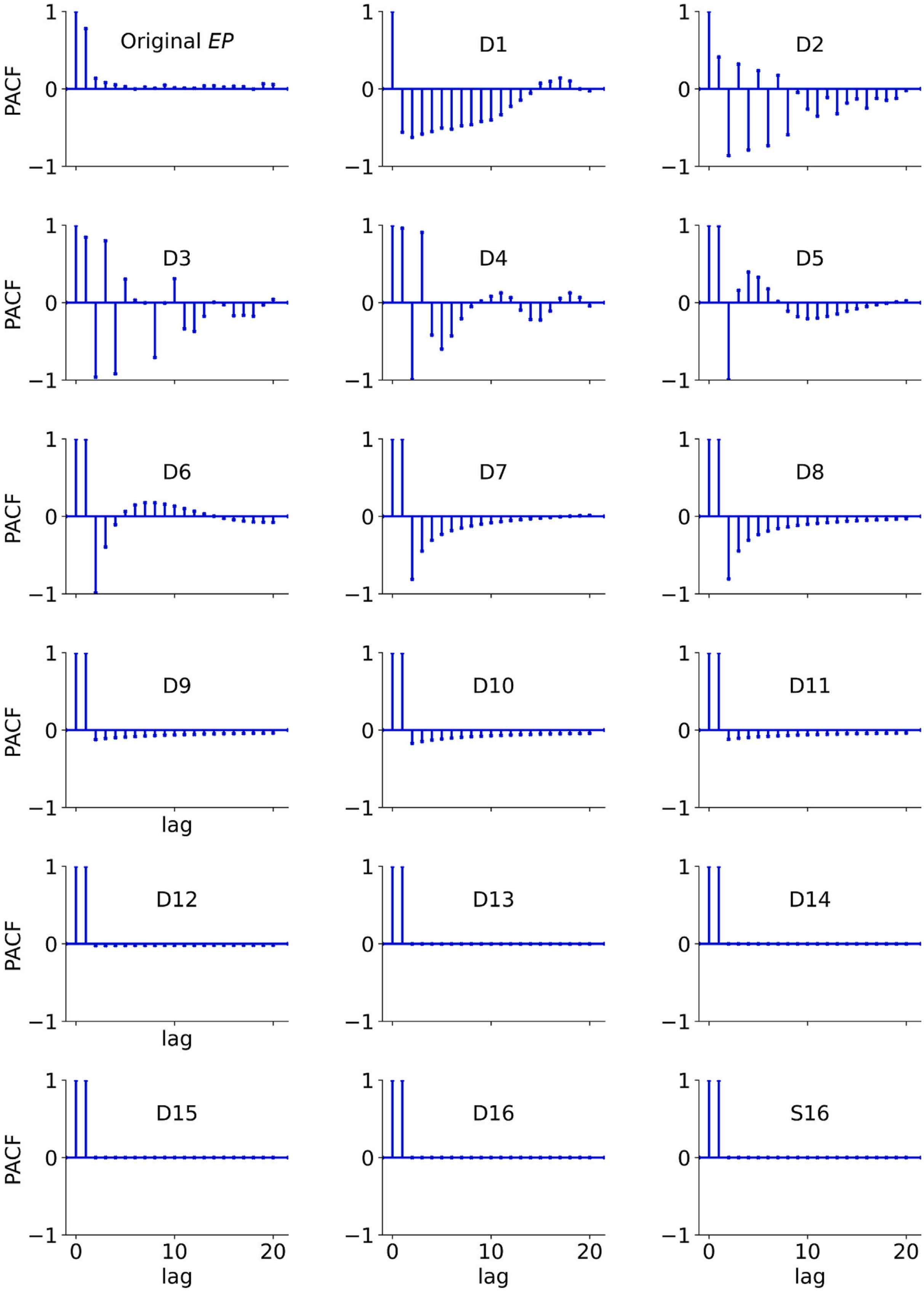


Fig. 9. Statistically significant Partial Autocorrelation Function (PACF) of detail and scaling coefficients for developing the proposed MoDWT-CRVFL model for $DS1$.

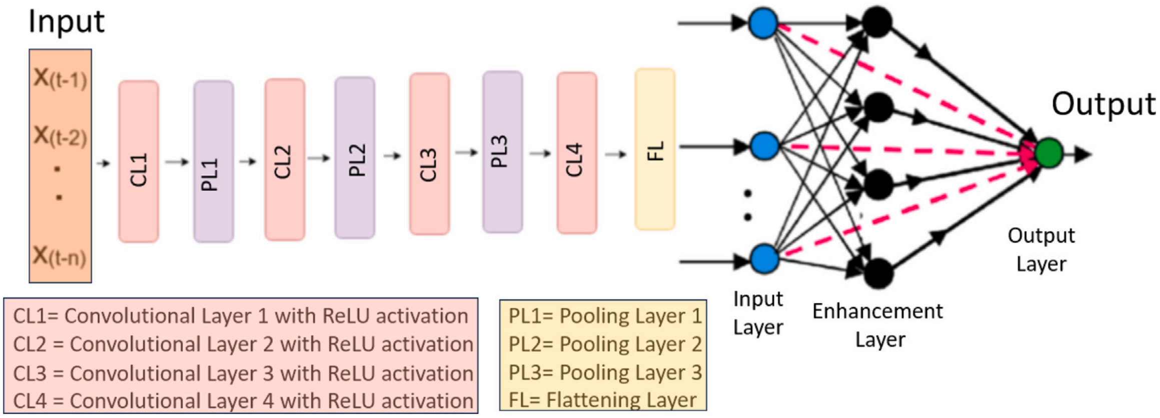


Fig. 10. An overview of CRVFL model for *EP* prediction. Note that this has the Convolutional layer, Pooling Layer, Flattening layer and the RVFL network. The bottom figure shows the CNN model parameter with each layers.

Table 4
The energy to Shannon entropy (R_{ES}) and the energy to Renyi entropy ratio (R_{ER}) test conducted after the MoDWT-based decomposition procedure applied in the electricity price data series using a different wavelet base.

Base wavelet	R_{ES}	R_{ER}	Base wavelet	R_{ES}	R_{ER}
db2	6920.7	7439.6	*sym9*	6886.8	7375.8
sym4	6912.3	7424.3	*db6*	6869.7	7351.8
coif2	6909.6	7422.1	*dmey*	6867.2	7346.6
fk6	6905	7414.8	*fk14*	6863.9	7336.6
sym6	6903.8	7408.5	*fk18*	6848.7	7309.9
sym5	6901.1	7398.2	*db8*	6846.9	7312.8
coif3	6900.4	7404.3	*db10*	6835.1	7290.2
sym8	6896.8	7396.7	*fk22*	6833.9	7287.4
coif4	6893.8	7392.7	*db12*	6817.6	7265.6
db4	6891.6	7383.8	*db14*	6807.8	7247
sym10	6891.1	7387.4	*db16*	6795.7	7229.9
fk8	6889.2	7388.8	*db18*	6787.8	7216
coif5	6888.9	7384.4	*db20*	6781.5	7207.2

Note: *coifN*, *dbN*, *fkN*, *symN*, and *demy* are Coiflets, Daubechies, Fejér-Korovkin, Symlets and the discrete approximation of Meyer wavelet,respectively. The index number corresponds to the coefficient count N of a wavelet.

of a data-driven model, and variations in their configurations can significantly impact the model's performance.

Consequently, identifying the optimal combination of hyperparameters is a fundamental aspect of constructing an effective data-driven model, particularly when dealing with hybrid deep learning models that possess numerous hyperparameters. Bayesian optimization (BO) stands out as an iterative algorithm widely employed in deep learning models [100]. In contrast to conventional grid search and random search methods, BO selects the next hyperparameter configuration based on the outcomes of previously tested configurations, enabling the discovery of superior hyperparameter combinations with fewer iterations, which is especially valuable when computational resources and time are limited.

This study has utilized the BO algorithm, integrated into the Optuna software [101] framework in Python 3.11, to conduct hyperparameter optimization for all standalone and hybrid models.

Table 7 presents the key hyperparameters and their corresponding search spaces for the CRVFL model. Furthermore, in this study, two callback functions were employed to prevent the model from over-fitting: Early Stopping (ES) and Reduce on Loss Plateau Decay ($ReduceLROnPlateau$) [102], configured with specific parameters. These parameters include patience set to 5, a factor of 0.5, and a minimum learning rate of $\min_{lr} = 1 \times 10^{-3}$. The Factor parameter serves as a multiplier that reduces the learning rate as $lr = lr \times factor$. The patience parameter denotes the number of epochs without improvement before the learning rate reduction is initiated. The ES callback is applied for early stopping of the model, while the $ReduceLROnPlateau$ callback adjusts the learning rate downward if a loss metric (Mean Square Error) fails to show improvement during validation.

Within the hyperparameter optimization process, the choice for the optimization algorithm is Adaptive Moment Estimation ($Adam$), employing a consistent learning rate (lr) of $1{\times}10^{-3}$, The exponential decay rate for the first moment estimates ($\beta_1 = 0.9$) and The exponential decay rate for the second moment estimates ($\beta_2 = 0.9999$), accompanied by an epsilon (ϵ) value of 1×10^{-8}. $Adam$ stands out as an efficient optimization algorithm, characterized by modest memory demands, resilience to gradient rescaling, and excellent suitability for managing extensive datasets [91].

Finally, after acquiring the optimal parameters through hyperparameter optimization, the model made its ultimate predictions on the test data. Subsequently, the seventeen levels of predicted output, comprising 16 detail coefficients and 1 scaling coefficient, are processed through iMoDWT (Inverse MoDWT) to yield the predicted *EP* series. It is noteworthy that the *EP* series obtained through CRVFL is in normalized form, and a de-normalization step is employed to recover the original *EP* series.

3.3. Benchmark model development

To evaluate thoroughly the proposed MoDWT-CRVFL model have utilized a comprehensive set of eleven different predictive models: six standalone models (MLP, XGB, LSTM, RF, DNN, Bi-LSTM) and five models using MoDWT decomposition (MoDWT-MLP, MoDWT-XGB, MoDWT-LSTM, MoDWT-RF, MoDWT-DNN) whose parameters are optimized using Bayesian method, as shown in Table 7.

All phases of the training process, encompassing hyperparameter optimization and testing for both the proposed model and benchmark models, were conducted on a workstation equipped with an Intel *i*7 4200U CPU operating at 1.6 GHz, complemented by 32 GB of RAM. The implementation of the algorithm was carried out using Python with a TensorFlow backend.

3.4. Performance criteria

To evaluate the performance of the proposed MoDWT-CRVFL model against the other benchmark models, this study has utilized standard metrics, as indicated by Eqs. (15)–(26), to measure prediction accuracy [103].

Mean Absolute Error (MAE):

$$MAE(\text{AUD/MWh}) = \frac{1}{N}\sum_{i=1}^{N}|EP^{p} - EP^{a}|, \tag{15}$$

Table 5

Input parameters corresponding to each decomposed series within the datasets $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$.

Dataset	Input parameters corresponding to each decomposed series
$DS1$	$D_{1,2}(x_{t-1}),\ D_3(x_{t-1}, x_{t-2}, x_{t-3}),\ D_{4,5,14,16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}),$ $D_{6-11,13,15}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}),\ D_{12}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}),$ $S_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$
DS2	$D_{1,2}(x_{t-1}),\ D_3(x_{t-1}, x_{t-2}, x_{t-3}),\ D_{4,5,14,15}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}),$ $D_{6-11,13,16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}),\ D_{12}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}),$ $S_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$
DS3	$D_{1-3}(x_{t-1}),\ D_{4,5,14-16}(x_{t-1}, x_{t-2}, x_{t-3}),\ D_{4,5,14-16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}),$ $D_{6-13}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}),\ S_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5})$
DS4	$D_{1,2}(x_{t-1}),\ D_3(x_{t-1}, x_{t-2}, x_{t-3}),\ D_{4,5,14,16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}),$ $D_{6-11,13,15}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}),\ D_{12}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}),$ $S_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6})$
DS5	$D_{1,2}(x_{t-1}),\ D_3(x_{t-1}, x_{t-2}, x_{t-3}),\ D_{4,5,14,15}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}),$ $D_{6-11,13}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}),\ D_{12}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}),$ $D_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7}, x_{t-8}),$ $S_{16}(x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7})$

Table 6

CNN architecture of the CRVFL model.

Layer (type)	Output shape	Params.
CL1 (Conv1d)	(10,10,100)	6100
PL1 (MaxPooling1D)	(10,5,100)	0
CL2 (Conv1d)	(10,5,80)	64 000
PL2 (MaxPooling1D)	(10,2,80)	0
CL3 (Conv1d)	(10,2,15)	7215
PL2 (MaxPooling1D)	(10,1,15)	0
CL4 (Conv1d)	(15)	1365
FL (Flatten)	(10,15)	0

Root Mean Square Error ($RMSE$):

$$RMSE(\text{AUD/MWh}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(EP^p - EP^a)^2}, \tag{16}$$

Symmetric Mean Absolute Percentage Error ($sMAPE$):

$$sMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|EP^a - EP^p|}{(|EP^a| + |EP^p|)/2}, \tag{17}$$

Legates and McCabe Index (I_{LM}):

$$I_{LM} = 1 - \frac{\sum_{i=1}^{N}|EP^p - EP^a|}{\sum_{i=1}^{N}|EP^a - \langle EP^a\rangle|}, \tag{18}$$

Nash–Sutcliffe Index (I_{NS}):

$$I_{NS} = 1 - \frac{\sum_{i=1}^{N}(EP^a - EP^p)^2}{\sum_{i=1}^{N}(EP^a - \langle EP^a\rangle)^2}, \tag{19}$$

Wilmott's Index (I_{WI}):

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

Absolute Percentage Bias (APB):

$$APB(\%) = \left|\frac{\sum_{i=1}^{n}(EP^a - EP^p)}{\sum_{i=1}^{n}EP^a}\right| \cdot 100, \tag{21}$$

Coefficient of Determination (R^2):

$$R^2 = \left(\frac{\sum_{i=1}^{N}(EP^a - \langle EP^a\rangle)(EP^p - \langle EP^p\rangle)}{\sqrt{\sum_{i=1}^{n}(EP^a - \langle EP^a\rangle)^2}\sqrt{\sum_{i=1}^{n}(EP^p - \langle EP^p\rangle)^2}}\right)^2, \tag{22}$$

Skill Score (SS) for model evaluations:

$$SS = 1 - \frac{RMSE(p,x)}{RMSE(pr,x)} \tag{23}$$

Kling–Gupta Efficiency:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\langle EP^p\rangle}{\langle EP^a\rangle} - 1\right)^2 + \left(\frac{CV^p}{CV^a}\right)^2} \tag{24}$$

Theil's Inequality Coefficient:

$$TIC = \frac{\sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(EP^p - EP^a)^2}}{\left(\sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(EP^a)^2} + \sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(EP^p)^2}\right)} \tag{25}$$

and RMSE Ratio:

$$RMSE_r = \frac{RMSE_1}{RMSE_2} \tag{26}$$

Notably, the values of EP^a and EP^p represent the actual and the predicted half-hourly *EP* while $\langle EP^a\rangle$ and $\langle EP^p\rangle$ represent the actual and predicted mean *EP*, N = number of tested data points, *p* stands for the model prediction, *x* for the observation, *pr* for perfect prediction (persistence), and *r* for the reference prediction, and CV is the Coefficient of Variation.$RMSE_1$ and $RMSE_2$ are objective and benchmark model $RMSE$ respectively. The details of physical explanations of the metrics used in evaluating the proposed MoDWT-CRVFL model are as follows:

- The R^2 metric falls within the [0, 1] range, while MAE and $RMSE$ are expressed in absolute units of *EP* (AUD/MWh) ranging from [0, +∞]. In this context, where 0 signifies a perfect model and +∞ indicates a poorly performing one, R^2 quantifies the proportion of variance explained by the model, whereas MAE solely quantifies error magnitude without considering its direction, and $RMSE$ gauges the average error magnitude while considering deviations from the actual values.
- The range of the I_{WI} metric lies between 0 and 1, representing an enhancement over the MAE and $RMSE$ metrics. I_{WI} effectively detects both additive and proportional disparities in means and variances between observed and simulated data.
- The I_{NS} metric spans from [−∞, 1] and is used to evaluate the relative magnitude of residual variance in comparison to measured variance, where a score of −∞ denotes the poorest fit, while 1 signifies a perfectly matched model.
- I_{LM} falls within the range of [0, 1]. It serves as a more resilient metric when compared to both I_{NS} and I_{WI}, specifically designed to address and surpass their inherent limitations.
- The model achieving the lowest Symmetric Mean Absolute Percentage Error ($sMAPE$) is regarded as the best performer. $sMAPE$ is a balanced metric that eliminates the division-by-zero problem.

Table 7
Hyperparameters search range using Bayesian optimization method.

Predictive models	Model hyperparameters	Hyperparameter selection
Convolution Neural Network Integrated with Random Vector Functional Link (CRVFL)	Filter 1 (CNN)	(‘Filter1’, range(50,120,5))
	Filter 2 (CNN)	(‘Filter2’, range(50,100,5))
	Filter 3 (CNN)	(‘Filter3’, range(20,80,5))
	Filter 4 (CNN)	(‘Filter4’, range(20,50,5))
	Epochs (CNN)	[1000]
	Activation function	[ReLU]
	Solver	[‘Adam’]
	Batch Size	(‘Batch_Size’, range(50,1500,200))
	Enhancement nodes	(‘n_nodes’, range(5,40,2))
	Regularization parameter	(regular_para’, range(0.5,2,0.5))
Deep Neural Network (DNN)	Hiddenneuron 1	(‘Units1’, range(50,120,5))
	Hiddenneuron 2	(‘Units2’, range(50,100,5))
	Hiddenneuron 3	(‘Units3’, range(50,80,5))
	Batch Size	(‘Batch_Size’, range(50,1500,50))
	Solver	[‘Adam’]
	Epochs	[1000]
Random Forest Regression (RF)	The maximum depth of the tree.	(‘max_depth’, range(1,20,1))
	The number of trees in the forest.	(‘n_estimators’, range(5,100,2))
	Minimum number of samples to split an internal node	(‘min_samples_split’, range(2,100,1))
	The number of features to consider when looking for the best split.	[‘auto’, ‘sqrt’, ‘log2’]
Long Short Term Memory Network (LSTM) and Bidirectional LSTM (Bi-LSTM)	LSTM cell 1	(‘Units 1’, range(50,100,5))
	LSTM cell 2	(‘Units 2’, range(50,80,5))
	Activation function	[ReLU]
	Epochs	[1000]
	Drop rate	(‘drop_rate’, range(0,0.5,0.1))
	Batch Size	(‘Batch_Size’, range(50,1500,200))
eXtreme Gradient Boosting (XGB)	Booster Type	gbtree’
	Step size shrinkage used in update to prevents overfitting.	(‘eta’, range(0.1,0.9,0.1))
	The maximum depth of the tree.	(‘max_depth’, range(1,20,1))
	The number of trees in the forest.	(‘n_estimators’, range(5,100,2))
Multi Layer Preceptron (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’]

Note: ‘gbtree’ = Gradient Boosted Trees, ReLU = Rectified Linear Unit Activation Function, and Adam = Adaptive Moment Estimation.

In contrast, the conventional Mean Absolute Percentage Error ($MAPE$) metric tends to exaggerate errors when the true value is near zero, whereas $sMAPE$ circumvents this issue.

- The Absolute Percentage Bias (APB) quantifies the error in predicted values as a percentage in relation to the observed values. A lower APB value, approaching zero, signifies better accuracy of the model. The ideal APB value is precisely zero.
- A negative Skill Score (SS) implies a prediction that is less accurate than the persistence model, while a positive SS signifies an improvement over the persistence model, with the extent of enhancement directly correlated to SS, wherein higher scores indicate more significant improvements. It is important to emphasize that the persistence model is a straightforward forecasting method that utilizes the previous time step’s value as the prediction for the next time step.
- The Kling–Gupta Efficiency (KGE) addresses several limitations of the Nash–Sutcliffe Efficiency (I_{NS}) and is increasingly favoured for model evaluation. Much like I_{NS} where $KGE = 1$ denotes perfect agreement between simulations and observations, a KGE value less than 0 indicates that the mean of observations provides more accurate estimates than the simulations.
- The Theil’s Inequality Coefficient (TIC) relies on observed bias, variance, and covariance, and its values range between zero for perfect prediction and one for very poor prediction performance. The closer the TIC approaches zero, the more accurate the forecasting performance.
- The $RMSE$ ratio ($RMSE_r$) is employed to evaluate models, and if a specific A model demonstrates superior performance compared to the B model It is compared to, then the RMSE ratio of A/B should be less than 1.

Choosing a model based on a single statistical metric is somewhat naive as each model has its unique merits and limitations. Thus, this study have adopted Global Performance Index (GPI) amalgamating multiple metrics for a holistic assessment of the proposed MoDWT-CRVFL mode, where a higher magnitude of the GPI corresponds to a greater accuracy. Instead of examining individual metrics, GPI ranges from 0 to 1, assigning equal weights to various statistical metrics [104]:

$$GPI_i = \sum_{j=1}^{n} \xi_j \left(\tilde{I}_i - I_{ij} \right), \tag{27}$$

where $\xi_j = -1$ for Pearson’s Correlation Coefficient and $\xi_j = 1$ for all the other indicators. $\tilde{I}_i$ is the median of scaled values of indicator j and I_{ij} is the scaled value of indicator for model.

The interpretation of GPI is that when this metric falls below the median, a larger disparity between that value and the median

Table 8
An evaluation of the proposed MoDWT-CRVFL model for half-hourly *EP* prediction for five dataset: $DS1$ (Winter); $DS2$ (Autumn); $DS3$ (Spring); $DS4$ (Summer); and $DS5$ (Yearly 2022) in terms of R^2 = Coefficient of Determination, $RMSE$: *AUD/MWh* = Root Mean Square Error; MAE:*AUD/MWh* = Mean Absolute Error.

		Decomposition based model						Standalone models					
Dataset	*Metrics*	MoDWT-CRVFL Objective Model	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
DS1	R^2	**0.998**	0.997	0.993	0.995	0.993	0.990	0.848	0.847	0.842	0.841	0.840	0.835
	RMSE	**3.895**	4.405	4.459	4.681	5.614	6.478	22.342	22.422	22.722	22.789	22.957	23.263
	MAE	**2.240**	3.078	2.818	2.380	2.965	3.480	11.880	11.896	12.044	12.046	12.093	12.862
DS2	R^2	**0.998**	0.995	0.994	0.955	0.956	0.974	0.943	0.944	0.940	0.935	0.937	0.941
	RMSE	**10.919**	16.242	22.770	48.385	47.749	33.627	49.049	48.839	50.004	52.054	51.310	49.678
	MAE	**7.782**	9.012	17.100	29.352	28.936	18.360	26.895	28.094	27.998	27.796	27.300	28.764
DS3	R^2	**0.999**	0.998	0.993	0.986	0.986	0.980	0.958	0.956	0.958	0.953	0.952	0.958
	RMSE	**10.924**	17.284	12.899	44.034	42.401	43.205	63.209	64.295	63.323	67.744	67.083	63.792
	MAE	**7.891**	11.485	9.188	25.777	24.710	27.446	41.657	42.488	42.432	43.902	43.032	42.832
DS4	R^2	**0.998**	0.990	0.996	0.994	0.991	0.996	0.879	0.881	0.878	0.876	0.878	0.877
	RMSE	**6.778**	14.243	8.851	11.095	13.837	8.400	43.269	43.258	43.631	44.363	44.055	43.892
	MAE	**4.611**	10.618	6.528	7.066	9.449	5.738	27.934	27.569	28.331	29.016	29.143	28.820
DS5	R^2	**0.999**	0.993	0.984	0.983	0.980	0.984	0.964	0.964	0.963	0.958	0.958	0.961
	RMSE	**8.617**	37.163	39.652	32.773	35.927	32.520	47.517	47.738	48.038	50.858	50.898	49.115
	MAE	**6.030**	22.495	24.611	18.808	20.546	18.903	27.831	27.807	28.688	28.769	28.415	29.935

of all other models implies that the model is more accurate than its counterparts. Conversely, when it surpasses the median, a greater deviation from the median suggests that the model is less accurate than other models.

In the testing phase, this study explored the statistical agreement between the actual and predicted half-hourly *EP* using the Diebold–Mariano (DM) statistic test. The DM statistic is defined as follows:

$$S_{DM} = \frac{\bar{g}}{\sqrt{\left(\widehat{V}_g/N\right)}}, \tag{28}$$

where

$$\bar{g} = \left(\sum_{t=1}^{N} g_t\right)/N,\ g_t = \left(x_t - \hat{x}_{te,t}\right)^2 - \left(x_t - \hat{x}_{re,t}\right)^2, \tag{29}$$

and

$$\widehat{V}_g = \gamma_0 + 2\sum_{t=1}^{\infty} \gamma_t\,, \quad \left(\gamma_t = \text{cov}\left(g_{t+1}, g_t\right)\right), \tag{30}$$

where γ_0 = variance of g_t, $\hat{x}_{te,t}$ and $\hat{x}_{re,t}$ represent the predicted values of x_t calculated using the tested method te and reference method re, respectively, in period t. N is the number of observations in testing dataset.

It is important to note that DM statistic evaluates the significance of different models [105] to determine whether the expected forecast accuracy is uniform across models. In this task, $RMSE$ acts as a loss function, with null hypothesis positing that the $RMSE$ of the tested model (te) is not lower than that of the reference model (re).

4. Result and discussions

This study now evaluates the performance of the proposed MoDWT-CRVFL model for half-hourly *EP* prediction. Initially, a preliminary assessment of all models based on the R^2, as well as non-standardized metrics such as $RMSE$ and MAE is conducted. As shown in Table 8, the proposed MoDWT-CRVFL model has a remarkable performance, achieving the highest R^2 ranging from 0.998 to 0.999, as well as the lowest values for $RMSE$ (ranging from 3.895 to 10.924 AUD/MWh) and MAE (ranging from 2.240 to 7.782 AUD/MWh) across all five datasets.

It is worth noting that R^2 for all five datasets were relatively similar in respect to the proposed MoDWT-CRVFL model and the other decomposition-based models. In contrast, the $RMSE$ and MAE values for all the other decomposition-based models were consistently higher compared to those of the MoDWT-CRVFL model.

Similarly, when comparing the standalone models (Bi-LSTM, LSTM, DNN, RF, XGB, and MLP) with both the proposed MoDWT-CRVFL model and the other decomposition-based models, it is evident that the $RMSE$ and MAE are notably higher for standalone models. This observation suggests that, particularly for *EP* series characterized by strong volatility, the decomposition of *EP* series using MoDWT is highly advantageous in reducing the data noise and the sequence frequency that perhaps leads to a more accurate prediction.

As a result, the $RMSE$ and MAE metrics exhibit substantial reductions, with decreases of 83% and 83% for $DS1$, 78% and 73% for $DS2$, 83% and 82% for $DS3$, 85% and 84% for $DS4$, and 85% and 84% for $DS5$, respectively. Furthermore, the R^2 values of the proposed MoDWT-CRVFL model show noteworthy improvements, increasing by 20%, 6%, 4%, 14%, and 4% for $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$ datasets, respectively when compared with standalone models.

The quantitative assessment of model performance, as quantified by Willmott's Index (I_{WI}), the Nash–Sutcliffe Index (I_{NS}), and Legates-McCabe's Index (I_{LM}), revealed a significant enhancement with the CRVFL model based on MoDWT. This improvement outperformed the performance of other deep learning models relying on MoDWT decomposition, namely MoDWT-LSTM and MoDWT-DNN, as well as the performance of individual standalone models, see refer to Table 9, for half-hourly *EP* prediction.

When comparing the proposed MoDWT-CRVFL model to decomposition-based deep learning models, it was noted that the smallest increase in the I_{LM} value, approximately 3.5%, occurred for the $DS4$ dataset, while a substantial increase of approximately 21.2% was observed for the $DS5$ dataset. Similarly, the I_{WI} value increased to around 0.996, representing an 8.7% increment, and I_{NS} reached 0.995, indicating an 8.5% increment for $DS4$. Moreover, when contrasting the standalone models (Bi-LSTM, LSTM, DNN, RF, XGB, and MLP) with the proposed MoDWT-CRVFL model, there were significant increases in the values of I_{WI}, I_{NS}, and I_{LM}.

Specifically, for the $DS1$ dataset, there were increments of 56.2%, 10.2%, and 169.2%, respectively. For $DS2$, there were increases of 15.4%, 26.1%, and 33.7%. For $DS3$, the values increased by 10.1%, 19.3%, and 44.3%. For $DS4$, there were substantial increments of 34.4%, 68.9%, and 121.2%. Finally, for $DS5$, the values increased by 12.5%, 16.7%, and 34.0%.

In Figs. 11 and B.15, scale-independent metrics, known as Symmetric Mean Absolute Percentage Error $(sMAPE)$, are showcased for

Table 9

An evaluation of the proposed MoDWT-CRVFL model for half-hourly *EP* prediction using the normalized, non-dimensional model evaluation metrics: I_{WI} = Willmott's Index; I_{NS} = Nash–Sutcliffe Index; I_{LM} = Legates and McCabe Index for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

		Decomposition based models						Standalone models					
Dataset	Metrics	MoDWT-CRVFL Objective Model	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
DS1	I_{WI}	**0.987**	0.983	0.982	0.982	0.974	0.965	0.592	0.591	0.609	0.604	0.619	0.632
	I_{NS}	**0.985**	0.981	0.981	0.979	0.969	0.959	0.517	0.513	0.500	0.497	0.490	0.476
	I_{LM}	**0.883**	0.839	0.853	0.876	0.845	0.818	0.379	0.378	0.371	0.371	0.368	0.328
DS2	I_{WI}	**0.993**	0.984	0.972	0.872	0.875	0.934	0.858	0.860	0.852	0.840	0.847	0.861
	I_{NS}	**0.990**	0.977	0.956	0.800	0.804	0.901	0.790	0.792	0.782	0.764	0.770	0.785
	I_{LM}	**0.915**	0.901	0.812	0.678	0.682	0.798	0.705	0.691	0.692	0.695	0.700	0.684
DS3	I_{WI}	**0.997**	0.993	0.996	0.954	0.957	0.954	0.889	0.883	0.892	0.877	0.879	0.905
	I_{NS}	**0.995**	0.988	0.993	0.922	0.927	0.924	0.838	0.832	0.837	0.813	0.817	0.834
	I_{LM}	**0.935**	0.906	0.924	0.788	0.797	0.774	0.658	0.651	0.651	0.639	0.646	0.648
DS4	I_{WI}	**0.994**	0.971	0.989	0.983	0.973	0.990	0.722	0.749	0.740	0.635	0.642	0.739
	I_{NS}	**0.990**	0.956	0.983	0.974	0.959	0.985	0.598	0.598	0.591	0.579	0.585	0.586
	I_{LM}	**0.905**	0.782	0.866	0.855	0.806	0.882	0.427	0.435	0.419	0.405	0.403	0.409
DS5	I_{WI}	**0.996**	0.935	0.920	0.944	0.933	0.945	0.880	0.880	0.875	0.867	0.872	0.886
	I_{NS}	**0.995**	0.918	0.905	0.934	0.921	0.935	0.863	0.861	0.859	0.842	0.842	0.853
	I_{LM}	**0.940**	0.776	0.754	0.812	0.795	0.811	0.722	0.723	0.714	0.713	0.717	0.701

Table 10

An evaluation of the proposed MoDWT-CRVFL model for half-hourly *EP* prediction using the normalized, evaluation metrics: $RRMSE$ = Relative Root Mean Square Error; and $RMAE$ = Relative Mean Absolute Error for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

		Decomposition based model						Standalone models					
Dataset	Metrics	MoDWT-CRVFL Objective Model	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
DS1	*RRMSE*	**5.09%**	5.76%	5.83%	6.12%	7.34%	8.47%	29.22%	29.33%	29.72%	29.80%	30.03%	30.43%
	RMAE	**3.62%**	5.23%	5.42%	3.86%	5.00%	5.58%	24.31%	24.12%	22.55%	22.78%	21.95%	21.94%
DS2	*RRMSE*	**5.86%**	8.71%	12.22%	25.96%	25.62%	18.04%	26.32%	26.20%	26.83%	27.93%	27.53%	26.65%
	RMAE	**5.76%**	6.11%	8.46%	18.12%	17.98%	14.62%	23.85%	23.93%	23.46%	23.96%	23.31%	22.20%
DS3	*RRMSE*	**3.73%**	5.90%	4.40%	15.03%	14.48%	14.75%	21.58%	21.95%	21.62%	23.13%	22.90%	21.78%
	RMAE	**5.47%**	7.85%	5.45%	12.23%	11.57%	18.05%	34.66%	36.21%	31.21%	33.84%	33.94%	26.10%
DS4	*RRMSE*	**4.54%**	9.54%	5.93%	7.43%	9.27%	5.63%	28.98%	28.97%	29.22%	29.71%	29.51%	29.40%
	RMAE	**5.99%**	13.84%	7.43%	9.78%	14.24%	8.30%	46.06%	41.08%	43.56%	61.87%	61.06%	40.91%
DS5	*RRMSE*	**4.78%**	20.60%	21.98%	18.17%	19.92%	18.03%	26.34%	26.47%	26.63%	28.20%	28.22%	27.23%
	RMAE	**6.44%**	11.72%	23.49%	20.42%	22.43%	19.93%	33.76%	32.56%	34.40%	31.81%	29.40%	29.33%

Table 11

An evaluation of the proposed MoDWT-CRVFL model for half-hourly *EP* predictions using the normalized, evaluation metrics: KGE = Kling Gupta Efficiency; APB = Absolute Percentage Bias; and TIC= Theil Inequality Coefficient for four different seasons and the yearly (i.e., 2022) prediction dataset with the best model indicated in blue.

		Decomposition based model						Standalone models					
Dataset	Metrics	MoDWT-CRFVL Objective Model	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
DS1	*KGE*	**0.972**	0.958	0.899	0.963	0.943	0.930	0.661	0.659	0.708	0.696	0.739	0.738
	APB	**2.930**	4.025	3.685	3.113	3.877	4.551	15.538	15.558	15.751	15.754	15.815	16.821
	TIC	**0.024**	0.026	0.027	0.028	0.034	0.039	0.137	0.137	0.139	0.139	0.139	0.142
DS2	*KGE*	**0.987**	0.933	0.856	0.761	0.767	0.929	0.895	0.908	0.856	0.882	0.875	0.886
	APB	**4.175**	4.836	9.175	15.748	15.525	9.851	14.430	15.073	15.022	14.914	14.647	15.433
	TIC	**0.025**	0.038	0.056	0.119	0.117	0.079	0.114	0.114	0.118	0.122	0.121	0.117
DS3	*KGE*	**0.983**	0.977	0.964	0.843	0.852	0.946	0.884	0.871	0.839	0.798	0.822	0.956
	APB	**2.694**	3.921	3.137	8.800	8.436	9.370	14.221	14.505	14.486	14.988	14.691	14.622
	TIC	**0.016**	0.026	0.020	0.069	0.066	0.065	0.095	0.097	0.097	0.105	0.103	0.096
DS4	*KGE*	**0.978**	0.927	0.922	0.944	0.915	0.962	0.757	0.807	0.767	0.664	0.667	0.802
	APB	**3.089**	7.111	4.372	4.732	6.329	3.843	18.709	18.465	18.975	19.433	19.518	19.302
	TIC	**0.021**	0.044	0.027	0.034	0.042	0.026	0.134	0.134	0.136	0.136	0.135	0.136
DS5	*KGE*	**0.969**	0.777	0.802	0.959	0.959	0.942	0.936	0.940	0.874	0.885	0.897	0.927
	APB	**3.343**	12.472	13.645	10.427	11.391	10.480	15.430	15.417	15.905	15.950	15.754	16.596
	TIC	**0.020**	0.090	0.095	0.074	0.082	0.074	0.107	0.108	0.111	0.117	0.117	0.113

half-hourly *EP* predictions during the test phase. These metrics evaluate both the proposed model and benchmark models across the $DS1$, $DS2$, $DS3$, $DS4$, and $DS5$ datasets.

When compared to the other models, the proposed MoDWT-CRVFL algorithm stands out with its remarkable overall performance, exhibiting a lower $sMAPE$ (3.04% for $DS1$, 5.55% for $DS2$,4.35% for $DS3$,4.82% for $DS4$, and 5.32% for $DS5$). The accuracy of the other decomposition-based models (Fig. 11) and standalone models (Fig. B.15) varied significantly, confirming that the proposed MoDWT-CRVFL model displayed greater potential for producing accurate half-hourly *EP* predictions.

During the testing phase, this study employed relative error metrics to assess model bias. The relative error values presented in Table 10

Table 12

The Diebold–Mariano (DM) test statistic for MoDWT-CRVFL *vs.* benchmark models. The column is compared with rows and if result is positive, the model in the row is superior to the column model and vice versa. The top-performing model is boldfaced (blue) with the objective model benchmarked against decomposition and standalone models for four seasons and the yearly 2022 dataset.

Table 11a: $DS1$ dataset.

Models	MoDWT-LSTM	MoDWT-XGB	MoDWT-MLP	MoDWT-RF	MoDWT-DNN	XGB	RF	DNN	MLP	LSTM	Bi-LSTM
MoDWT-CRVFL	**7.822**	**2.093**	**4.491**	**3.658**	**3.286**	**6.397**	**6.432**	**6.301**	**6.363**	**6.350**	**6.466**
MoDWT-LSTM		0.796	3.835	2.844	0.340	6.356	6.391	6.258	6.323	6.308	6.423
MoDWT-XGB			6.300	3.865	−0.753	6.497	6.544	6.395	6.459	6.456	6.576
MoDWT-MLP				−3.885	−4.045	6.478	6.527	6.366	6.435	6.432	6.556
MoDWT-RF					−2.792	6.526	6.569	6.416	6.483	6.483	6.608
MoDWT-DNN						6.363	6.402	6.259	6.325	6.314	6.432
XGB							−0.855	−1.028	1.185	−3.074	−2.829
RF								−0.230	1.493	−2.407	−2.447
DNN									2.752	−1.294	−1.283
MLP										−3.218	−2.706
LSTM											−0.537

Table 11b: $DS2$ dataset.

Models	MoDWT-LSTM	MoDWT-XGB	MoDWT-MLP	MoDWT-RF	MoDWT-DNN	XGB	RF	DNN	MLP	LSTM	Bi-LSTM
MoDWT-CRVFL	**9.601**	**7.095**	**9.515**	**12.729**	**9.467**	**12.866**	**12.309**	**11.769**	**11.770**	**11.630**	**10.763**
MoDWT-LSTM		−5.537	−8.246	−0.775	−8.185	12.491	11.914	11.305	11.356	11.146	10.340
MoDWT-XGB			−5.324	13.748	−4.132	13.195	12.585	12.028	12.004	11.890	10.950
MoDWT-MLP				11.955	2.206	12.805	12.253	11.712	11.715	11.581	10.695
MoDWT-RF					−10.123	12.776	12.195	11.636	11.632	11.497	10.571
MoDWT-DNN						12.699	12.150	11.616	11.613	11.471	10.606
XGB							1.395	−0.940	−0.349	−1.654	−1.605
RF								−1.994	−1.069	−2.803	−2.698
DNN									0.572	−1.741	−0.941
MLP										−1.525	−1.581
LSTM											0.028

Table 11c: $DS3$ dataset.

Models	MoDWT-LSTM	MoDWT-XGB	MoDWT-MLP	MoDWT-RF	MoDWT-DNN	XGB	RF	DNN	MLP	LSTM	Bi-LSTM
MoDWT-CRVFL	**3.667**	**7.084**	**6.883**	**7.117**	**9.224**	**7.172**	**6.836**	**6.348**	**7.125**	**8.330**	**7.072**
MoDWT-LSTM		7.073	7.394	7.124	4.925	7.530	7.150	6.615	7.516	8.783	7.463
MoDWT-XGB			−4.810	−3.967	−6.475	0.975	1.136	0.505	0.439	0.183	0.228
MoDWT-MLP				4.789	−4.360	6.791	6.201	5.526	6.564	8.251	6.414
MoDWT-RF					−6.473	1.213	1.357	0.717	0.671	0.451	0.458
MoDWT-DNN						6.203	5.914	5.422	6.096	7.259	6.017
XGB							1.605	−2.542	−4.448	−1.862	−5.268
RF								−5.160	−3.948	−1.962	−4.656
DNN									−0.549	−0.718	−1.505
MLP										−0.667	−2.007
LSTM											0.180

Table 11d $DS4$ dataset.

Models	MoDWT-LSTM	MoDWT-XGB	MoDWT-MLP	MoDWT-RF	MoDWT-DNN	XGB	RF	DNN	MLP	LSTM	Bi-LSTM
MoDWT-CRVFL	**10.756**	**4.242**	**12.447**	**4.066**	**7.249**	**12.974**	**13.006**	**12.550**	**12.864**	**13.098**	**13.109**
MoDWT-LSTM		3.860	12.334	3.664	−8.513	12.987	13.001	12.547	12.885	13.124	13.134
MoDWT-XGB			−0.177	−6.242	−4.157	6.315	6.683	4.928	4.902	5.044	4.789
MoDWT-MLP				−0.175	−12.351	11.780	11.796	11.216	11.382	11.585	11.603
MoDWT-RF					−3.976	6.741	7.109	5.344	5.301	5.447	5.198
MoDWT-DNN						12.944	12.983	12.515	12.829	13.059	13.073
XGB							1.742	−5.804	−3.316	−3.712	−4.862
RF								−5.921	−3.593	−4.109	−5.113
DNN									0.609	1.787	−0.201
MLP										0.738	−0.952
LSTM											−4.346

Table 11e: $DS5$ dataset.

Models	MoDWT-LSTM	MoDWT-XGB	MoDWT-MLP	MoDWT-RF	MoDWT-DNN	XGB	RF	DNN	MLP	LSTM	Bi-LSTM
MoDWT-CRVFL	**10.391**	**14.878**	**15.323**	**14.634**	**14.680**	**15.012**	**14.803**	**14.371**	**15.657**	**15.122**	**15.358**
MoDWT-LSTM		−3.121	−3.303	−0.909	2.761	9.877	9.581	7.533	8.807	7.627	7.611
MoDWT-XGB			−2.536	11.063	10.217	14.264	13.926	13.246	15.247	14.388	14.761
MoDWT-MLP				9.225	10.429	13.923	13.547	12.778	14.699	13.904	14.178
MoDWT-RF					6.021	14.342	14.096	13.529	15.849	14.693	15.285
MoDWT-DNN						12.354	11.752	9.175	11.423	9.575	9.626
XGB							−0.219	−7.134	−5.053	−7.457	−7.475
RF								−7.928	−5.105	−6.834	−7.498
DNN									4.434	−0.933	−1.722
MLP										−4.762	−6.480
LSTM											−1.086

demonstrate that the proposed MoDWT-CRVFL model consistently exhibited the lowest $RRMSE$ and $RMAE$ across all datasets.

These relative metrics, $RRMSE$ and $RMAE$, are straightforward to interpret, categorizing models into “Excellent” ($RRMSE/MAE \leq$ 10%), “Good” (10% $\geq RRMSE/MAE \leq$ 20%), “Fair” (20% $\geq RRMSE/MAE \leq$ 30%), and “Poor” ($RRMSE/MAE \geq$ 30%) based on percentage criteria. As a result, the MoDWT-CRVFL model received an “Excellent” rating for all datasets. Conversely, the performance of other

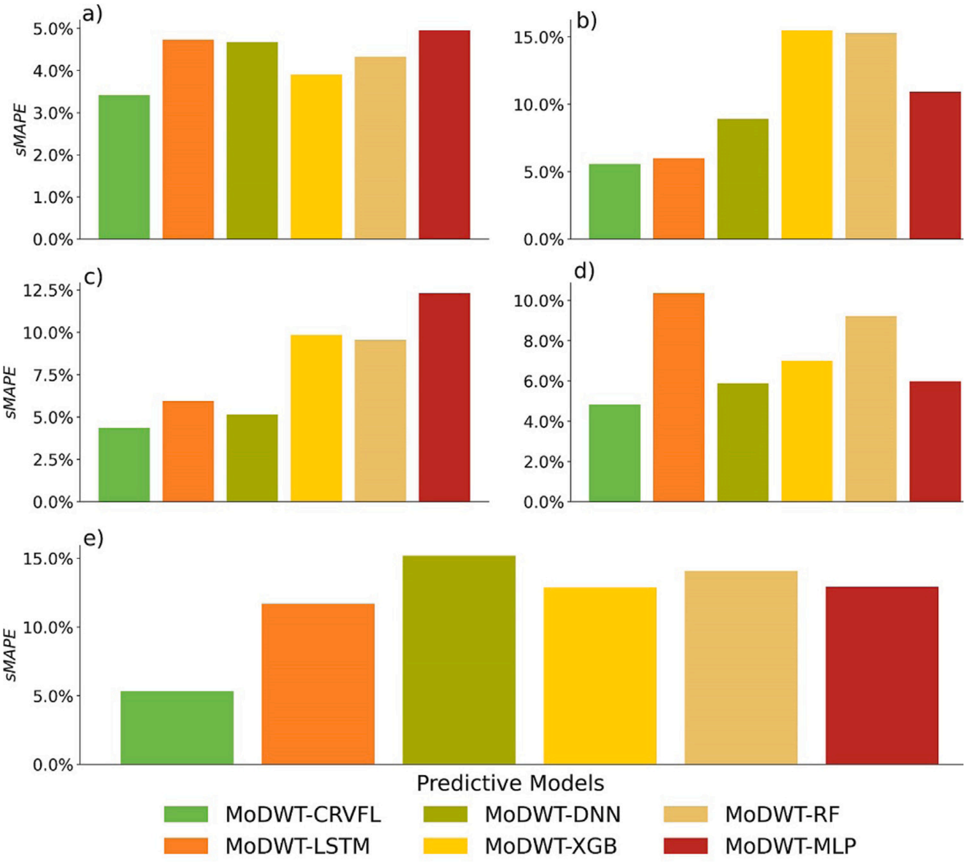


Fig. 11. The Symmetric Mean Absolute Percentage Error (sMAPE) for the testing period for: (a) $DS1$ (b) $DS2$, (c) $DS3$, (d) $DS4$, and (e) $DS5$ datasets for decomposition based models.

decomposition-based models varied. For example, in the $DS5$ dataset, the $RRMSE$ | $RMAE$ values were 20.60% | 11.72%, 21.98% | 23.49%, 18.17% | 20.42%, 19.2% | 22.44%, and 18.03% | 19.93% for MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, and MoDWT-MLP, respectively.

The proposed MoDWT-CRVFL model demonstrated the most significant reduction in $RRMSE$ values across all five datasets. Compared to MoDWT-LSTM, it achieved a reduction of 11.6% and 30.9% for $DS1$, a reduction of 52.0% and 31.9% for $DS2$ when compared to MoDWT-DNN, a reduction of 75.2% and 55.3% for $DS3$ when compared to MoDWT-XGB, a reduction of 51.0% and 58.0% for $DS4$ when compared to MoDWT-RF and a reduction of 73.5% and 67.7% for $DS5$ when compared to MoDWT-MLP in terms of the $RRMSE$ and $RMAE$, respectively.

Similarly, for the standalone models, both the $RRMSE$ and the $RMAE$ exceeded by 20% across all five datasets. When compared to MoDWT-CRVFL, the value of $RRMSE$ and $RMAE$ actually increased by 82.6% and 85.1% for Bi-LSTM ($DS1$), 77.6% and 75.9% for LSTM ($DS2$), 82.7% and 82.5% for DNN ($DS3$), and 84.7% and 90.3% for XGB ($DS5$), 83.1% and 78.1% for RF ($DS4$), and 76.4% and 78.1% for MLP ($DS5$).

The model's Skill Score (SS) (see Fig. B.18) as well as the predicted *vs.* actual electricity prices (see Fig. B.19) also serve as a means to evaluate its effectiveness of the proposed MoDWT-CRVFL model in comparison with the benchmark-persistence model. In addition, the results in Figs. 12 and B.16 also show the boxplots to visually assess the performance of the proposed MoDWT-CRVFL model in conjunction with other decomposition-based and standalone models. These figures incorporate markers to identify outliers within the absolute Prediction Error ($|PE|$) for the testing data, while also representing the upper quartile, median, and lower quartile. The distributions of $|PE|$, as illustrated by the proposed MoDWT-CRVFL model across all datasets, consistently demonstrate notably smaller quartile ranges.

It is noteworthy, that the decomposition-based and standalone models exhibited a distorted spread characterized by overestimated upper quartiles. Therefore, the proposed MoDWT-CRVFL model exhibited superior predictive performance, as evidenced by the box plots. This finding is further substantiated by the previously assessed performance metrics such as R^2, MAE, $RMSE$, I_{LM}, I_{NS}, I_{WI}, $RMAE$, $RRMSE$ and $sMAPE$.

An assessment of the distribution of prediction errors ($|PE|$) was also conducted to gauge the effectiveness of the proposed MoDWT-CRVFL model. Figs. 13 and B.17 portrays histograms representing the probability distribution of $|PE|$, computed within error brackets of 5-step increments. A more detailed analysis of the probability distribution of $|PE|$ across all models further underscores the robustness and appropriateness of the proposed model.

Specifically, the proposed MoDWT-CRVFL model exhibited the highest percentage of $|PE|$ within the first bin ($0 \leq |PE| \leq 5.0$), registering at ≈57%, surpassing the $|PE|$ percentages of ≈45%, ≈38%, ≈42%, ≈38%, and ≈41% observed for MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, and MoDWT-MLP, respectively. Additionally, in this study, the evaluation of the proposed MoDWT-CRVFL model extended beyond the use of the Nash–Sutcliffe Coefficient (I_{NS}). It encompassed an analysis of the three components of the I_{NS}, which include Correlation, Bias, and the Ratio of Variances (Coefficients of Variation). These components were scrutinized in Table 11 to provide a balanced assessment using the Kling–Gupta Efficiency (KGE).

The KGE metrics for the proposed model consistently outperformed both decomposition-based and standalone models. For instance, the KGE value was ≈ 0.972 for MoDWT-CRVFL, surpassing values of ≈0.958, ≈0.899, ≈0.963, ≈0.943, ≈0.930, ≈0.661, ≈0.708, ≈0.696,

Table 13
Evaluating the proposed MoDWT-CRVFL model with Root Mean Square Error ratio ($RMSE_r$) in testing phase *vs.* against decomposition and standalone models for four different seasons and 2022 dataset.

Table 12a: $DS1$ dataset.

Models	MoDWT-CRVFL Objective Model	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
MoDWT-CRVFL Objective Model	**1.000**	**1.131**	**1.145**	**1.202**	**1.441**	**1.663**	**5.736**	**5.757**	**5.834**	**5.851**	**5.895**	**5.973**
MoDWT-LSTM	**0.884**	1.000	1.012	1.063	1.274	1.471	5.072	5.090	5.158	5.173	5.211	5.281
MoDWT-DNN	**0.873**	0.988	1.000	1.050	1.259	1.453	5.010	5.028	5.095	5.110	5.148	5.217
MoDWT-XGB	**0.832**	0.941	0.953	1.000	1.199	1.384	4.773	4.790	4.854	4.868	4.904	4.969
MoDWT-RF	**0.694**	0.785	0.794	0.834	1.000	1.154	3.980	3.994	4.048	4.059	4.090	4.144
MoDWT-MLP	**0.601**	0.680	0.688	0.723	0.867		3.449	3.461	3.507	3.518	3.544	3.591
Bi-LSTM	**0.174**	0.197	0.200	0.210	0.251	0.290	1.000	1.004	1.017	1.020	1.028	1.041
LSTM	**0.174**	0.196	0.199	0.209	0.250	0.289	0.996		1.013	1.016	1.024	1.037
DNN	**0.171**	0.194	0.196	0.206	0.247	0.285	0.983	0.987	1.000	1.003	1.010	1.024
RF	**0.171**	0.193	0.196	0.205	0.246	0.284	0.980	0.984	0.997	1.000	1.007	1.021
XGB	**0.170**	0.192	0.194	0.204	0.245	0.282	0.973	0.977	0.990	0.993	1.000	1.013

Table 12b: $DS2$ dataset.

Models	MoDWT-CRVFL	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
MoDWT-CRVFL	**1.000**	**1.487**	**2.085**	**4.431**	**4.373**	**3.080**	**4.492**	**4.473**	**4.580**	**4.767**	**4.699**	**4.550**
MoDWT-LSTM	**0.672**	1.000	1.402	2.979	2.940	2.070	3.020	3.007	3.079	3.205	3.159	3.059
MoDWT-DNN	**0.480**	0.713	1.000	2.125	2.097	1.477	2.154	2.145	2.196	2.286	2.253	2.182
MoDWT-XGB	**0.226**	0.336	0.471	1.000	0.987	0.695	1.014	1.009	1.033	1.076	1.060	1.027
MoDWT-RF	**0.229**	0.340	0.477	1.013	1.000	0.704	1.027	1.023	1.047	1.090	1.075	1.040
MoDWT-MLP	**0.325**	0.483	0.677	1.439	1.420	1.000	1.459	1.452	1.487	1.548	1.526	1.477
Bi-LSTM	**0.223**	0.331	0.464	0.986	0.973	0.686	1.000	0.996	1.019	1.061	1.046	1.013
LSTM	**0.224**	0.333	0.466	0.991	0.978	0.689	1.004	1.000	1.024	1.066	1.051	1.017
DNN	**0.218**	0.325	0.455	0.968	0.955	0.672	0.981	0.977	1.000	1.041	1.026	0.993
RF	**0.210**	0.312	0.437	0.930	0.917	0.646	0.942	0.938	0.961	1.000	0.986	0.954
XGB	**0.213**	0.317	0.444	0.943	0.931	0.655	0.956	0.952	0.975	1.015	1.000	0.968

Table 12c: $DS3$ dataset.

Models	MoDWT-CRVFL	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
MoDWT-CRVFL	**1.000**	**1.582**	**1.181**	**4.031**	**3.881**	**3.955**	**5.786**	**5.886**	**5.797**	**6.201**	**6.141**	**5.840**
MoDWT-LSTM	**0.632**	1.000	0.746	2.548	2.453	2.500	3.657	3.720	3.664	3.920	3.881	3.691
MoDWT-DNN	**0.847**	1.340	1.000	3.414	3.287	3.350	4.900	4.985	4.909	5.252	5.201	4.946
MoDWT-XGB	**0.248**	0.393	0.293	1.000	0.963	0.981	1.435	1.460	1.438	1.538	1.523	1.449
MoDWT-RF	**0.258**	0.408	0.304	1.038	1.000	1.019	1.491	1.516	1.493	1.598	1.582	1.504
MoDWT-MLP	**0.253**	0.400	0.299	1.019	0.981	1.000	1.463	1.488	1.466	1.568	1.553	1.477
Bi-LSTM	**0.173**	0.273	0.204	0.697	0.671	0.684	1.000	1.017	1.002	1.072	1.061	1.009
LSTM	**0.170**	0.269	0.201	0.685	0.659	0.672	0.983	1.000	0.985	1.054	1.043	0.992
DNN	**0.173**	0.273	0.204	0.695	0.670	0.682	0.998	1.015	1.000	1.070	1.059	1.007
RF	**0.161**	0.255	0.190	0.650	0.626	0.638	0.933	0.949	0.935	1.000	0.990	0.942
XGB	**0.163**	0.258	0.192	0.656	0.632	0.644	0.942	0.958	0.944	1.010	1.000	0.951

Table 12d $DS4$ dataset.

Models	MoDWT-CRVFL	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
MoDWT-CRVFL	**1.000**	**2.101**	**1.306**	**1.637**	**2.041**	**1.239**	**6.383**	**6.382**	**6.437**	**6.545**	**6.499**	**6.475**
MoDWT-LSTM	**0.476**	1.000	0.621	0.779	0.971	0.590	3.038	3.037	3.063	3.115	3.093	3.082
MoDWT-DNN	**0.766**	1.609	1.000	1.254	1.563	0.949	4.889	4.888	4.930	5.012	4.978	4.959
MoDWT-XGB	**0.611**	1.284	0.798	1.000	1.247	0.757	3.900	3.899	3.933	3.999	3.971	3.956
MoDWT-RF	**0.490**	1.029	0.640	0.802	1.000	0.607	3.127	3.126	3.153	3.206	3.184	3.172
MoDWT-MLP	**0.807**	1.696	1.054	1.321	1.647	1.000	5.151	5.150	5.194	5.281	5.245	5.225
Bi-LSTM	**0.157**	0.329	0.205	0.256	0.320	0.194	1.000	1.000	1.008	1.025	1.018	1.014
LSTM	**0.157**	0.329	0.205	0.256	0.320	0.194	1.000	1.000	1.009	1.026	1.018	1.015
DNN	**0.155**	0.326	0.203	0.254	0.317	0.193	0.992	0.991	1.000	1.017	1.010	1.006
RF	**0.153**	0.321	0.200	0.250	0.312	0.189	0.975	0.975	0.984	1.000	0.993	0.989
XGB	**0.154**	0.323	0.201	0.252	0.314	0.191	0.982	0.982	0.990	1.007	1.000	0.996

Table 12e: $DS5$ dataset.

Models	MoDWT-CRVFL	MoDWT-LSTM	MoDWT-DNN	MoDWT-XGB	MoDWT-RF	MoDWT-MLP	Bi-LSTM	LSTM	DNN	RF	XGB	MLP
MoDWT-CRVFL	**1.000**	**4.313**	**4.602**	**3.803**	**4.169**	**3.774**	**5.514**	**5.540**	**5.575**	**5.902**	**5.907**	**5.700**
MoDWT-LSTM	**0.232**	1.000	1.067	0.882	0.967	0.875	1.279	1.285	1.293	1.369	1.370	1.322
MoDWT-DNN	**0.217**	0.937	1.000	0.827	0.906	0.820	1.198	1.204	1.211	1.283	1.284	1.239
MoDWT-XGB	**0.263**	1.134	1.210	1.000	1.096	0.992	1.450	1.457	1.466	1.552	1.553	1.499
MoDWT-RF	**0.240**	1.034	1.104	0.912	1.000	0.905	1.323	1.329	1.337	1.416	1.417	1.367
MoDWT-MLP	**0.265**	1.143	1.219	1.008	1.105	1.000	1.461	1.468	1.477	1.564	1.565	1.510
Bi-LSTM	**0.181**	0.782	0.834	0.690	0.756	0.684	1.000	1.005	1.011	1.070	1.071	1.034
LSTM	**0.181**	0.778	0.831	0.687	0.753	0.681	0.995		1.006	1.065	1.066	1.029
DNN	**0.179**	0.774	0.825	0.682	0.748	0.677	0.989	0.994	1.000	1.059	1.060	1.022
RF	**0.169**	0.731	0.780	0.644	0.706	0.639	0.934	0.939	0.945	1.000	1.001	0.966
XGB	**0.169**	0.730	0.779	0.644	0.706	0.639	0.934	0.938	0.944	0.999	1.000	0.965

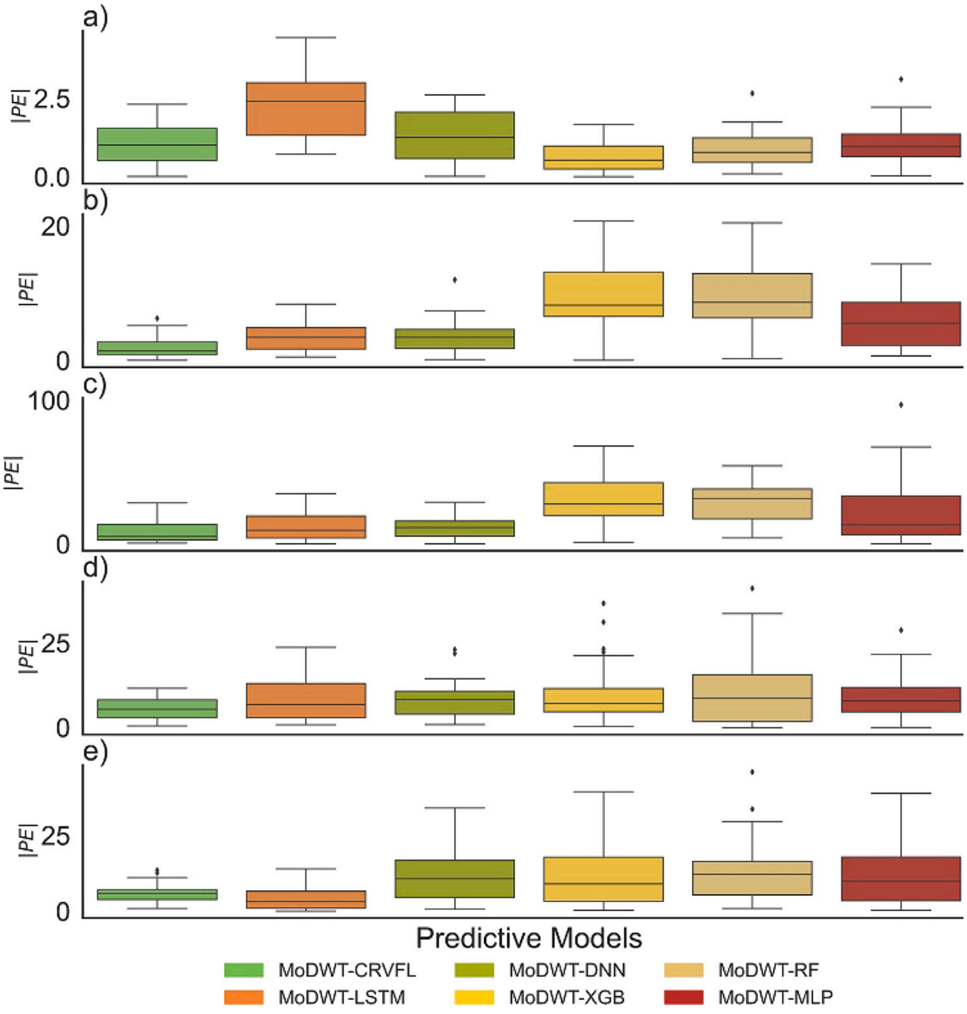


Fig. 12. Assessing the distribution of the Prediction Error $|PE|$ using a box plot analysis of the MoDWT-CRVFL model *vs.* several other decomposition based models.

≈0.739, and ≈0.738 for MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, MoDWT-MLP, Bi-LSTM, LSTM, DNN, RF, XGB, and MLP, respectively. Additionally, the Absolute Percentage Bias (APB) and Theil's inequality coefficient (TIC) were employed to validate the proposed model's half-hourly *EP* prediction capability.

As presented in Table 11, the error values for APB and TIC consistently favoured the proposed model over all benchmark models, affirming that the hybrid MoDWT-CRVFL model, utilizing MoDWT decomposition techniques, exhibited superior performance.

In summary, the efficacy of the MoDWT-CRVFL model was further corroborated through the use of KGE, APB, and TIC metrics. With a relatively high KGE score and comparatively low APB and TIC values, these results underscored the superior predictive performance of the MoDWT-CRVFL model, surpassing that of its counterparts.

While an extensive array of evaluation metrics and diagnostic plots were employed for comparing models, reliably ranking a large number of models based on these metrics can be challenging. Therefore, a more robust and comprehensive Global Performance Indicator (GPI) was introduced for this purpose.

In Fig. 14, the GPI chart illustrates the performance of the proposed MoDWT-CRVFL model in comparison to decomposition-based models. The results revealed that the proposed MoDWT-CRVFL model consistently outperformed other decomposition-based models.

For instance, concerning the $DS1$ dataset, the GPI was approximately 1.6122 for MoDWT-CRVFL, while it was ≈ 1.4651, ≈1.4624, ≈1.5226, ≈1.3649, and ≈1.2398 for MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, and MoDWT-MLP, respectively. Similarly, the GPI values for standalone models were ≈-0.05398, ≈-0.067715, ≈-0.085263, ≈-0.13501, ≈-0.10207, and ≈-0.070849 for Bi-LSTM, LSTM, DNN, RF, XGB, and MLP models. Therefore, this ranking based on the GPI unequivocally validates the superior accuracy and performance of the MoDWT-CRVFL model in half-hourly *EP* prediction.

While certain error indicators can indeed reflect disparities in prediction accuracy among the models, It is worth noting that these results may sometimes be misleading due to the influence of data characteristics on accuracy differences. Therefore, this study employed a Diebold–Mariano (DM) test to further quantify the distinctions in accuracy between the models. Analysing the outcomes presented in Table 12, several key observations can be made.

First and foremost, it is evident that the proposed MoDWT-CRVFL model consistently outperforms both the other decomposition-based models and the standalone models, as indicated by positive DM test results. Secondly, the standalone models exhibit inferior performance compared to all other decomposition-based models. Lastly, among the decomposition-based models, the MoDWT-LSTM model emerges as the top performer after the MoDWT-CRVFL model.

Finally, this study conducted a comparison based on the Root Mean Square Error ratio ($RMSE_r$) between the proposed MoDWT-CRVFL model and the benchmark models (Table 13). Upon examining the $RMSE_r$ values for the proposed MoDWT-CRVFL model in contrast to both decomposition-based and standalone models, it becomes evident that the MoDWT-CRVFL consistently outperforms the other models, with $RMSE_r$ consistently below 1.

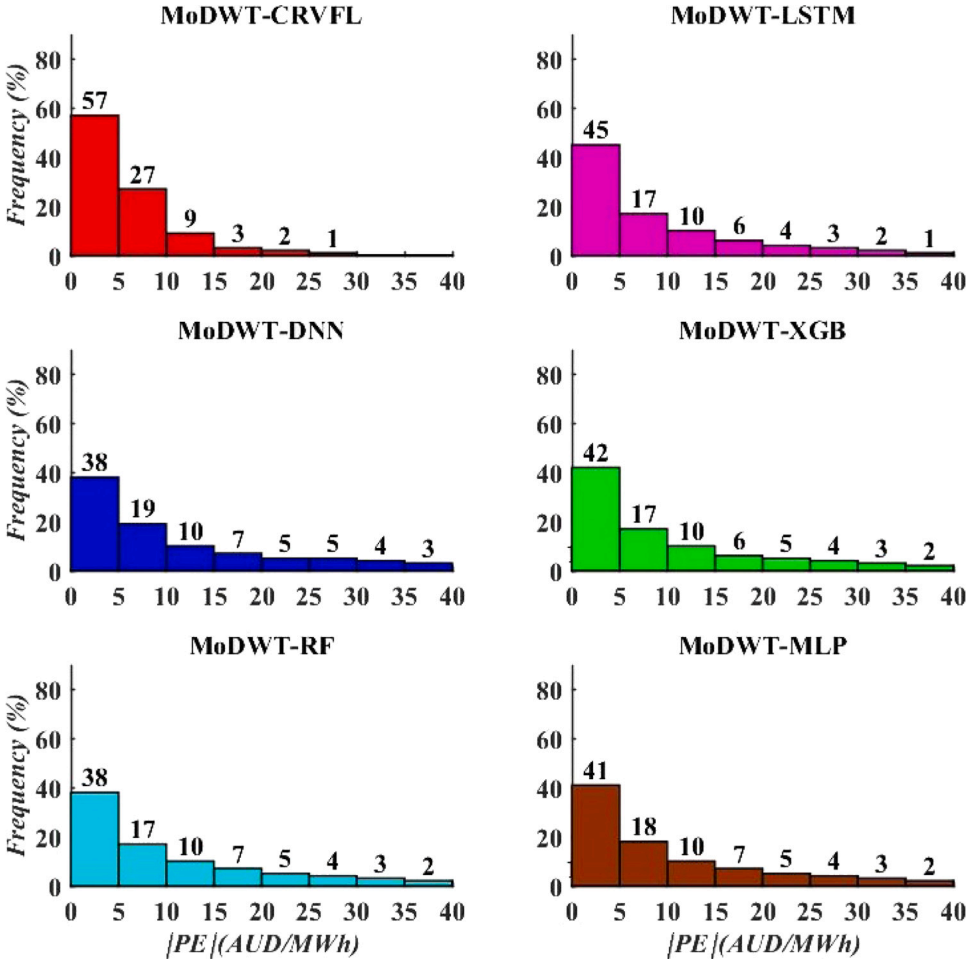


Fig. 13. Frequency of Absolute Prediction Error Analysis: Histograms of the proposed MoDWT-CRVFL model compared to 5 Other decomposition based models for half-hourly *EP* Prediction.

Additionally, among the decomposition-based deep learning models, MoDWT-LSTM and MoDWT-DNN emerge as the second and third best models, as indicated by $RMSE_r$ values consistently exceeding 0.4 across $DS1$, $DS2$, $DS3$ and $DS4$ datasets. However, for dataset $DS5$, the $RMSE_r$ values of decomposition based deep learning models fall within the range of 0.21–0.23.

4.1. Comparison of computational complexity

To illustrate the practical applicability of the proposed MoDWT-CRVFL model in real urban energy systems, this study have compiled and presented the computation times of the proposed model alongside five decomposition-based models and six standalone models in Table 14. Upon analysing these computation time results, it becomes apparent that the proposed model exhibits slightly lower computational efficiency compared to the other benchmark models. In this study, It is worth noting that the training time for standalone models (Bi-LSTM, LSTM, DNN, RF, XGB, and MLP) is relatively short, albeit at the expense of high prediction accuracy.

Conversely, for decomposition-based models, both training and testing are performed on all decomposed series, resulting in longer training times compared to standalone models. This is particularly true for models with complex structures and numerous parameters, which inevitably consume more time during both training and testing phases. However, It is crucial to highlight that the prediction accuracy of the other decomposition-based models (MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, and MoDWT-MLP) is inferior to that of the MoDWT-CRVFL model. Fortunately, despite this minor difference in computational efficiency, the proposed model consistently excels in prediction performance. The calculation time for the proposed model stands at 26.87 min, significantly below the one-hour mark in terms

Table 14
The computation time of the MoDWT-CRVFL vs. the benchmark models.

Models	Hyperparameter optimization time (min)	Training time (min)	Testing time (min)
MoDWT-CRVFL	161.23	44.79	26.87
MoDWT-LSTM	156.75	41.20	24.73
MoDWT-DNN	139.73	42.10	22.39
MoDWT-XGB	159.43	42.99	26.87
MoDWT-RF	131.67	28.66	25.08
MoDWT-MLP	87.78	29.56	22.39
Bi-LSTM	64.49	16.12	10.75
LSTM	62.70	15.23	12.54
DNN	51.95	14.33	13.44
RF	37.62	13.44	12.54
XGB	51.05	12.54	13.44
MLP	47.47	10.75	13.44

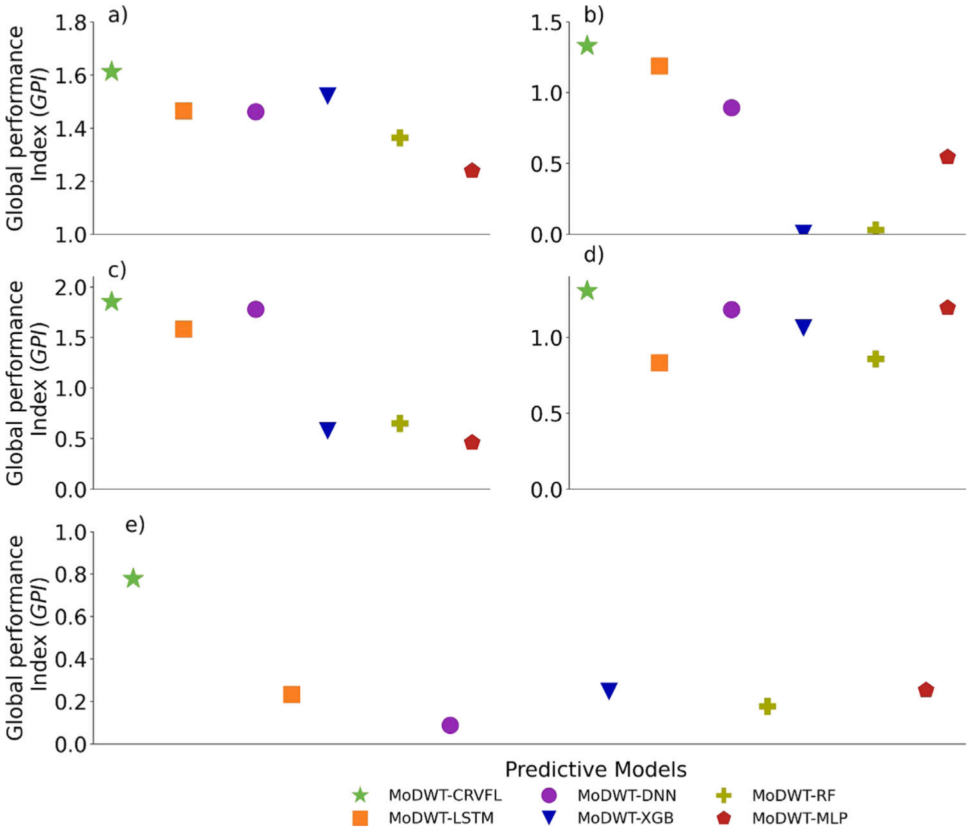


Fig. 14. Overall Global Performance Indicator (GPI) for (a) $DS1$ (b) $DS2$, (c) $DS3$, (d) $DS4$, and (e) $DS5$ datasets. Only the decomposition-based models are shown.

of the prediction time dimension. As a result, the proposed MoDWT-CRVFL model remains highly practical and effective for implementation in real urban energy systems.

4.2. Real application, limitations and future research work

The developed decomposition based prediction model holds substantial potential for widespread application within the Electric Power System and the broader electricity market. Notably, it offers an effective tool for all market participants, enabling producers and consumers to optimize production schedules and bidding strategies to maximize benefits. Simultaneously, market managers can utilize the predicted data to ensure the healthy and orderly operation of the power market, while also informing critical decisions related to transmission expansion and investment strategies.

While the developed decomposition-based prediction model demonstrates enhanced accuracy and stability in predicting future electricity prices by isolating distinct components, reducing noise, and converting non-stationary data into stationary series, these models are complex, computationally demanding, and pose a risk of overfitting. Their performance is heavily dependent on the quality of the initial decomposition, and integrating component predictions can be challenging, warranting attention in future research endeavours. This study primarily focuses on electricity price prediction, leaving room for the exploration of the following research directions:

(a) Expanding the input features beyond historical price data to include factors like electricity demand and temperature, enhancing the reliability and practicality of the prediction model.
(b) The decomposition strategy in data preprocessing, while beneficial for prediction performance, increases computational efficiency. Hence, future work should focus on optimizing this strategy.
(c) In comparison to point prediction, probabilistic interval prediction offers valuable insights for uncertainty analysis.

In respect to other future studies, providing probabilistic interval prediction results, alongside point-based price (or demand) predictions will deliver more comprehensive information to all energy market participants. A future study can also adopt error compensation strategies [2,106] to improve the performance of any price prediction model and build an integrated multi-head self-attention transformer model with local climate [103] to include their role in influencing seasonal and annual weather-event electricity price fluctuations.

5. Conclusions

In this paper we have introduced a novel hybrid model for Electricity Price (*EP*) prediction. Specifically, the MoDWT-CRVFL algorithm has been proposed, which amalgamates the MoDWT (Maximum Overlap Discrete Wavelet Transform) decomposition technique with a Bayesian optimized (BO) Convolutional Neural Network (CNN), integrated with a Random Vector Functional Link (RVFL). The proposed MoDWT-CRVFL model leverages the strengths of MoDWT for data decomposition and CRVFL (CNN+RVFL) for time series processing. The idea behind our approach is that it combines a robust time series decomposition capability, such as the MoDWT technique with the strong feature extraction capability of the CNN, and the powerful nonlinear fitting predictability of the Random Vector Functional Link (RVFL). This integration has been shown an excellent approach to process data information, to accurately predict short-term EP.

The results obtained in this research work showed that the use of the MoDWT algorithm effectively retains the valuable information from the original *EP* time series, and therefore eliminates the noise, and bolsters the data signal-to-noise ratios through a denoising effect. The optimization of the CRVFL hyperparameters using BO further enhances the proposed MoDWT-CRVFL model's ability to handle time series data. To evaluate the proposed MoDWT-CRVFL model's predictive performance and generalization capabilities, extensive experiments have been conducted using historical half-hourly *EP* data from New South Wales, Australia's electricity markets. The datasets were categorized into seasons (Summer: $DS1$, Autumn: $DS2$, Winter: $DS3$, and Spring: $DS4$) and also annually ($DS5$). Notably, in the volatile New South Wales electricity market, characterized by significant price fluctuations, the proposed model excels by accurately capturing price trends and effectively predicting price spikes.

The proposed MoDWT-CRVFL model's predictive performance has been systematically assessed by comparing it with eleven alternative benchmark models, encompassing both decomposition-based models (MoDWT-LSTM, MoDWT-DNN, MoDWT-XGB, MoDWT-RF, and MoDWT-MLP) and standalone models (Bi-LSTM, LSTM, DNN, RF, XGB, and MLP). Moreover, a dedicated performance evaluation module has been devised to assess the model's capabilities. The simulation results unequivocally demonstrated the superior performance of the proposed MoDWT-CRVFL model across all datasets ($DS1$, $DS2$, $DS3$, $DS4$, and $DS5$), attesting to its reliability in predicting electricity prices (*EP*).

To illustrate its exceptional performance of the proposed MoDWT-CRVFL model, this study now refers to specific results of two test cases of the Summer ($DS1$) and Autumn ($DS2$) seasons to conclude the following:

- For the case of $DS1$, where the proposed MoDWT-CRVFL model achieved an (R^2) of ≈0.996, (I_{WI}) of ≈0.987, (I_{NS}) of ≈0.985, (I_{LM}), of ≈0.883, $RMSE$ of ≈3.895 AUD/MWh, and an MAE of ≈2.240 AUD/MWh. Furthermore, the relative prediction error, $RRMSE$ registered a value of ≈5.094%, $RMAE$ of ≈3.616% whereas the normalized metric KGE was ≈0.972, APB was ≈2.930%, TIC was ≈0.02, and the $sMAPE$ value was ≈3.4%.
- For the case of $DS2$, the results generated by the proposed MoDWT-CRVFL model were also notably accurate, featuring an (R^2) of ≈0.997, (I_{WI}) of ≈0.993, (I_{NS}) of ≈0.990, (I_{LM}) of ≈0.915, $RMSE$ of ≈10.919 AUD/MWh, and MAE of ≈7.782 AUD/MWh. Likewise, the $RRMSE$ was ≈5.858%, $RMAE$ was ≈5.760%, KGE was ≈0.987, APB was ≈4.175%, TIC was ≈0.

Compared with all other decomposition-based and standalone models, the performance metrics stated above are impressively better to indicate the efficacy of the MoDWT-CRVFL model, and agrees with the results of other testing periods such as the Winter ($DS3$), Spring ($DS4$) and Yearly 2022 ($DS5$) that underpins the proposed model's reliability for half-hourly electricity price prediction. The Global Performance Indicator (GPI) ranking both the proposed and the benchmark models (Fig. 14) revealed that the proposed MODWT-CRVFL model consistently achieved highest GPI, which unequivocally justifies its competitive advantage for half-hourly *EP* predictions.

The comprehensive findings of this study indicate that, compared to benchmark models, the proposed MoDWT-CRVFL model excels not only in effectively managing the intricate characteristics inherent in *EP* series but also in delivering superior prediction and stability. The integration of MoDWT technique effectively addresses non-linearity and lack of smoothness in *EP* series, resulting in a significant enhancement of model's prediction accuracy, evidenced by its superior metrics contrasted with standalone models. These outcomes underscore the model's capacity to accurately predict stochastic and non-stationary *EP* series, thus offering invaluable support to market participants.

CRediT authorship contribution statement

Sujan Ghimire: Writing – original draft, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ravinesh C. Deo:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Conceptualization. **David Casillas-Pérez:** Writing – review & editing, Validation. **Ekta Sharma:** Writing – review & editing. **Sancho Salcedo-Sanz:** Writing – review & editing, Validation. **Prabal Datta Barua:** Writing – review & editing. **U. Rajendra Acharya:** Writing – review & editing.

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.

Data availability

Data were acquired from AEMO. (https://www.aemo.com.au/).

Acknowledgements

The authors thank data providers, all reviewers and Editor for their thoughtful comments, suggestions and review process. Partial support of this work was through a project PID2020-115454GB-C21 of the Spanish Ministry of Science and Innovation. Partial support of this work was through the LATENTIA project PID2022-140786NB-C31 of the Spanish Ministry of Science, Innovation and Universities (MICINNU).

Appendix A. Acronyms

Table A.15 provide the acronyms used in this study.

Appendix B. Additional results

In Appendix B, additional results regarding the performance of standalone models are shown using various visual analysis graphs. These results clearly ascertain the superior performance of the proposed MoDWT-CRVFL model against the benchmark (or standalone) models.

Fig. B.18 presents a polar plot illustrating the SS. Notably, the prediction skill score of the proposed model MoDWT-CRVFL outperformed all other models, underscoring its efficacy. For example, across all datasets, MoDWT-CRVFL achieved a SS of over 95% compared to the benchmark models. Among the decomposition-based deep learning models, MoDWT-LSTM and MoDWT-DNN demonstrated comparable results to the proposed model for $DS1$, $DS2$, $DS3$, and $DS4$ datasets. However, for $DS5$, their SS values were 49.3% and 42.54%, respectively. In contrast, standalone models exhibited SS scores ranging from 0.1% to 22.91%, with the RF model having the lowest SS of 0.1% for the $DS3$ dataset.

Figs. B.19 and B.20 display the actual and predicted *EP* generated by the proposed MoDWT-CRVFL model, alongside results from other decomposition-based and standalone models, during the testing period. To maintain brevity, this study have included plots for a 1-day prediction interval using the $DS1$ dataset. The figures (Figs. B.19 and B.20) clearly illustrate that the MoDWT-CRVFL model's predictions closely align with the actual *EP* values, surpassing the performance of the other models.

Appendix C. Initial values of model hyperparameters

In Appendix C, the initial values of the hyperparameters used for the development of the predictive model are presented in Tables C.16.

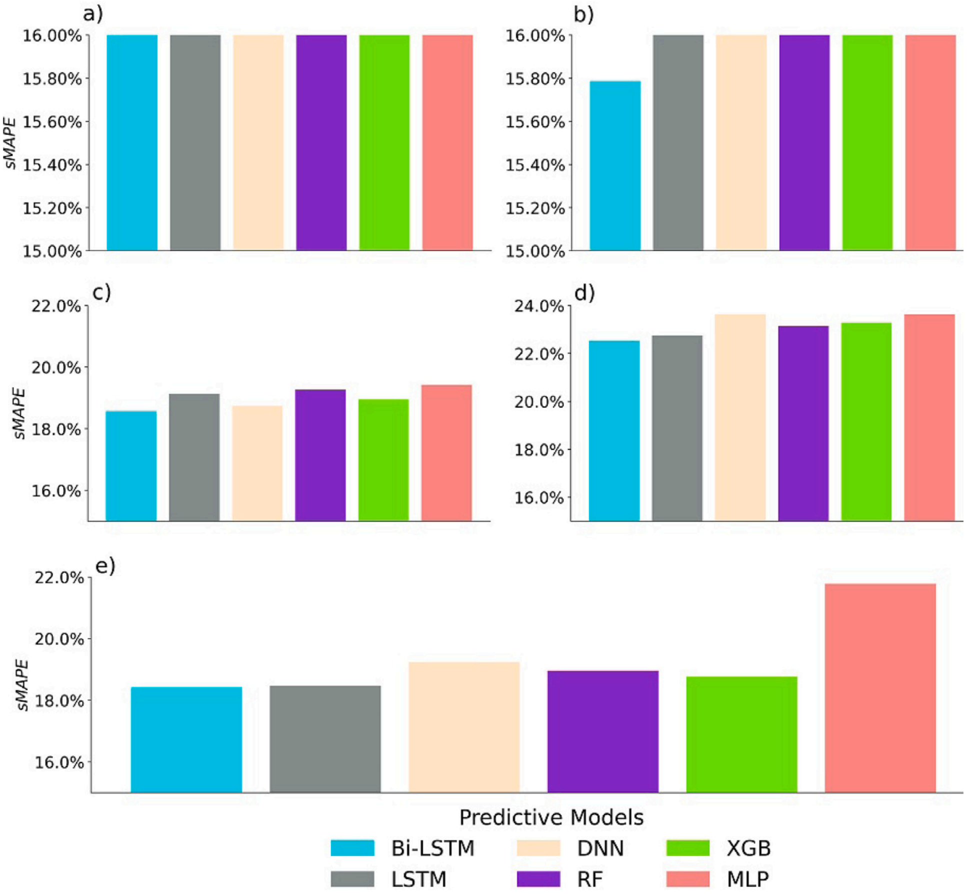


Fig. B.15. The Symmetric Mean Absolute Percentage Error (sMAPE) for the testing period in: (a) $DS1$ (b) $DS2$, (c) $DS3$, (d) $DS4$ and (e) $DS5$ datasets for standalone models.

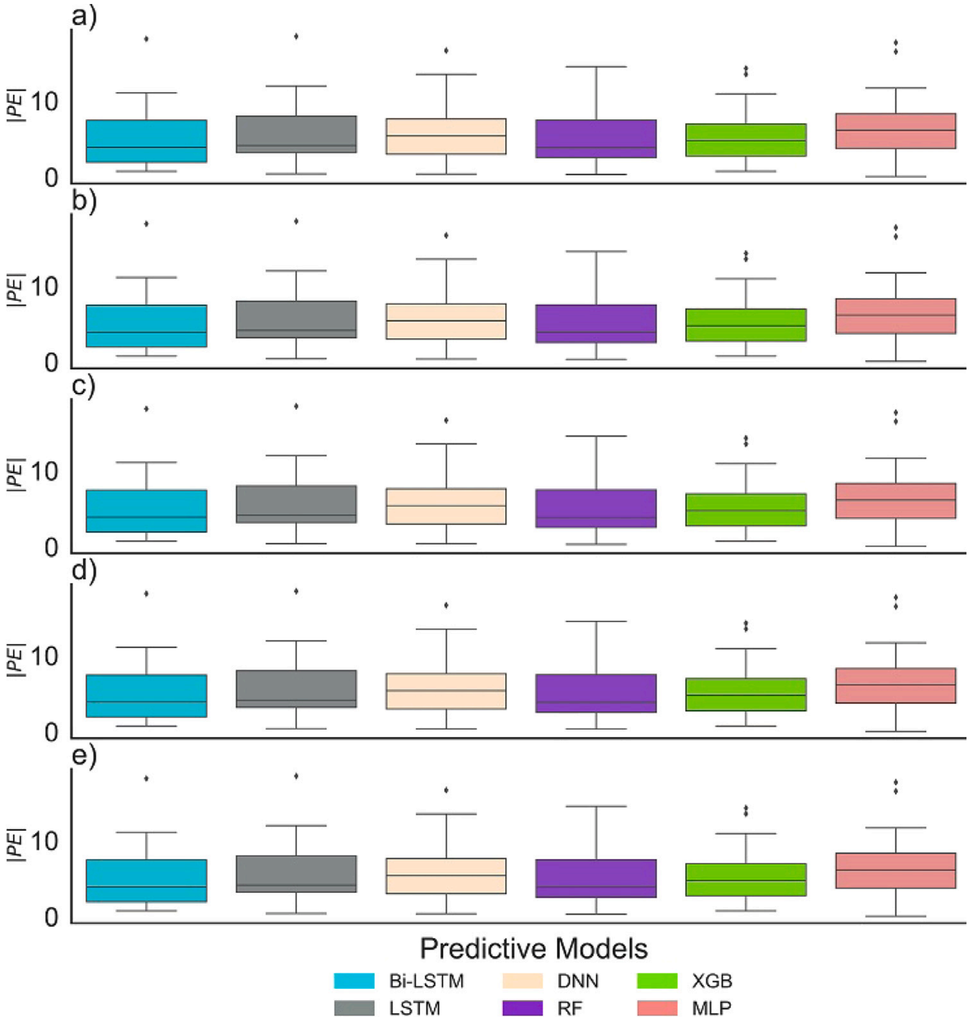


Fig. B.16. Assessing the distribution of the Prediction Error $|PE|$ using a box plot analysis of standalone models.

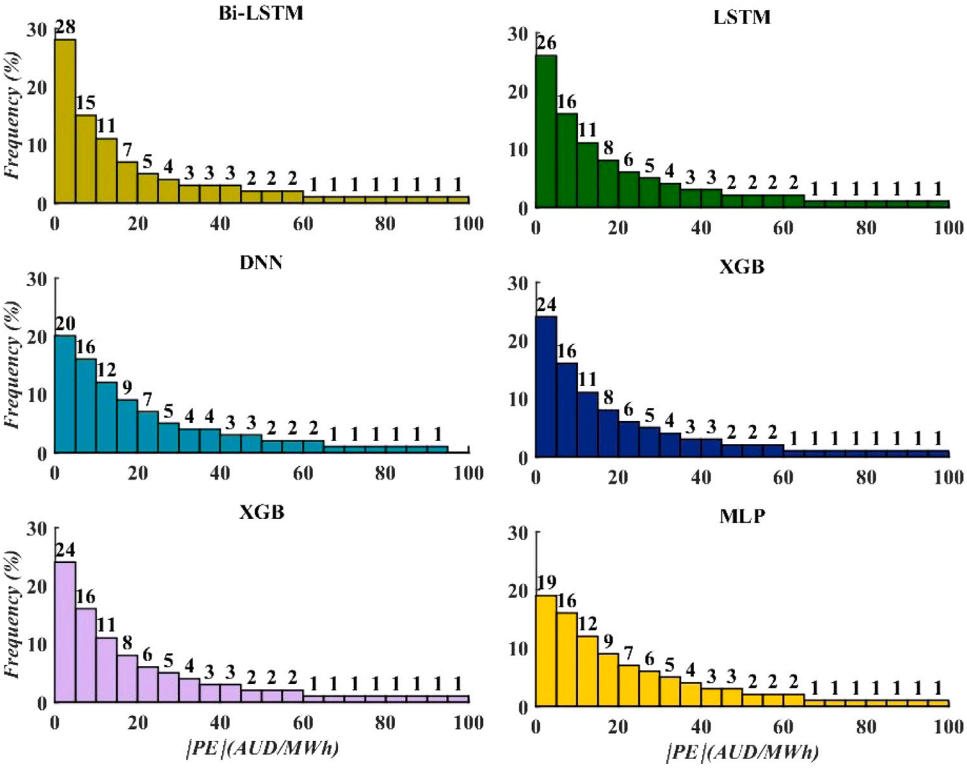


Fig. B.17. Frequency of Absolute Prediction Error Analysis: Histograms of the standalone models for half-hourly *EP* Prediction.

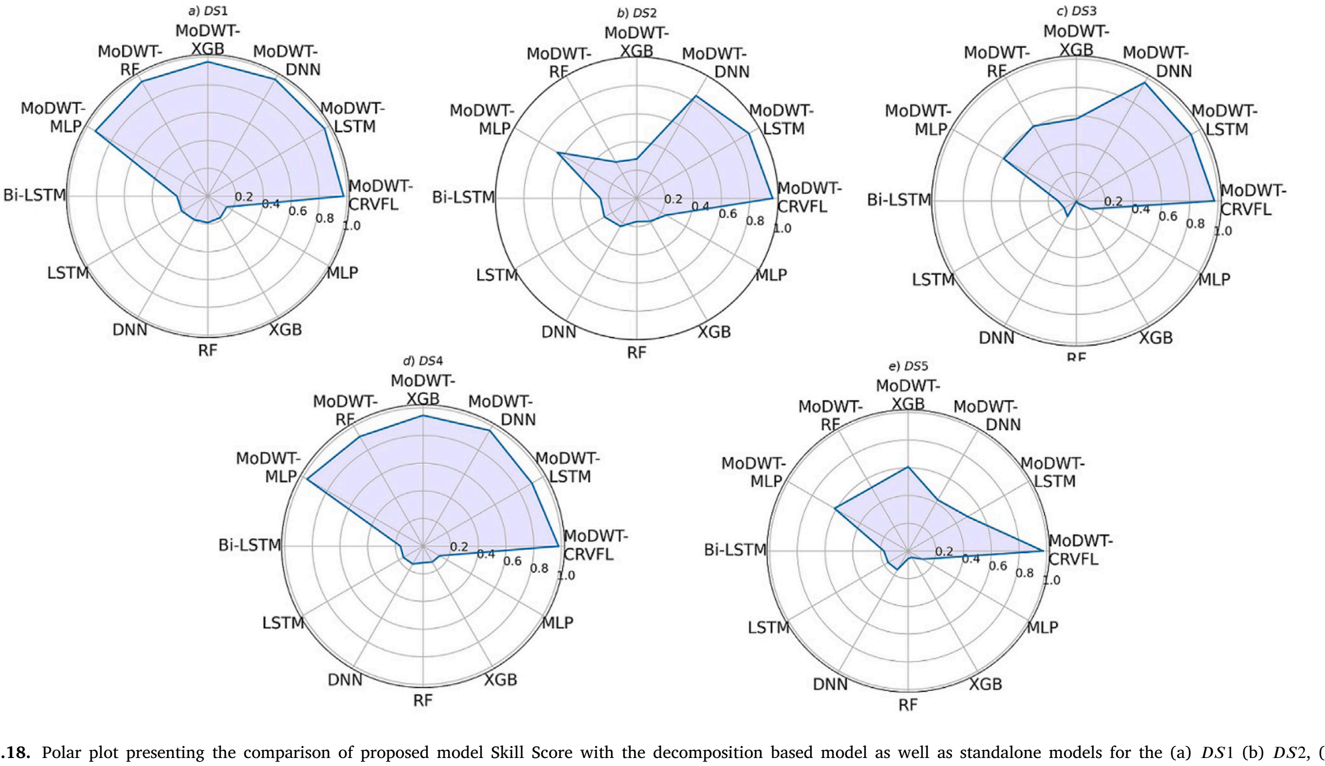


Fig. B.18. Polar plot presenting the comparison of proposed model Skill Score with the decomposition based model as well as standalone models for the (a) $DS1$ (b) $DS2$, (c) $DS3$, (d) $DS4$, and (e) $DS5$ datasets.

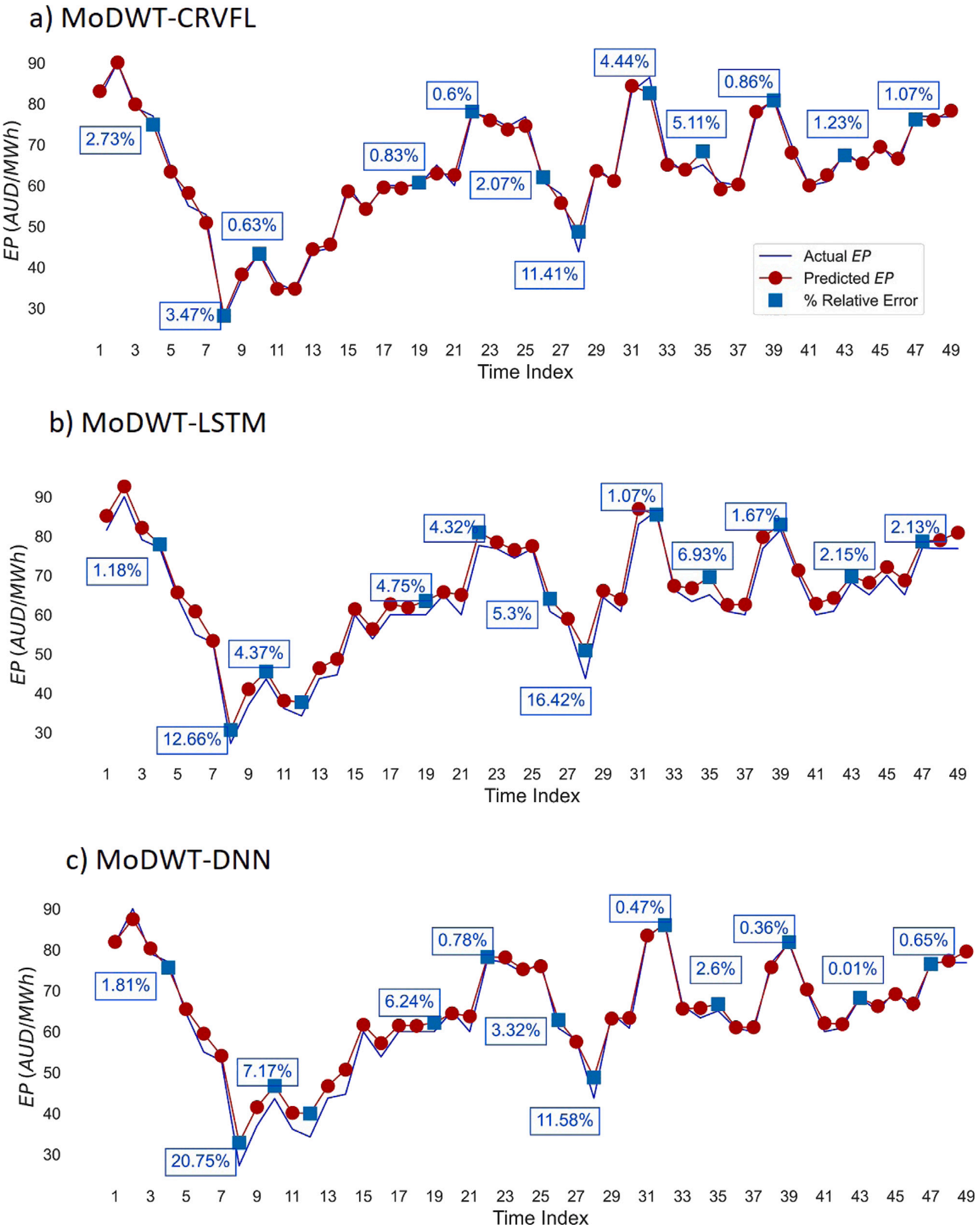
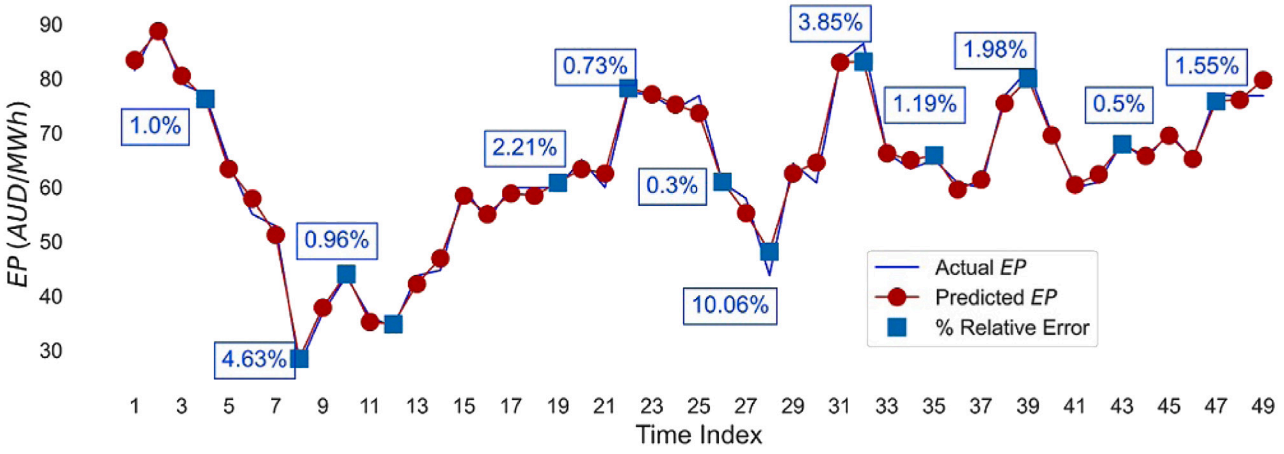
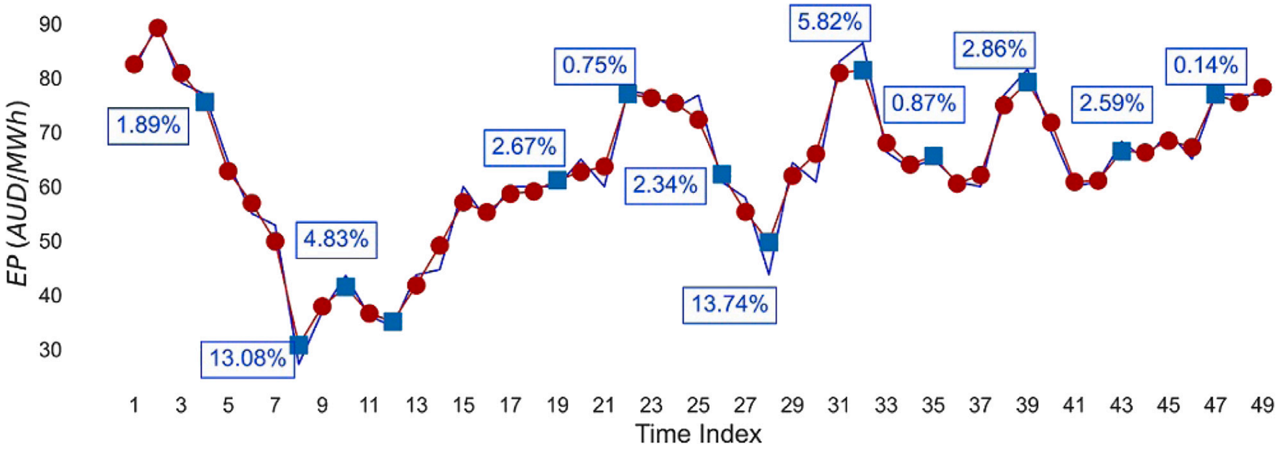


Fig. B.19. Comparative Evaluation of Predicted *vs.* Actual Electricity Prices (AUD/MWh) for the $DS1$ Dataset During a 2-Day Test Period Using a decomposition based models. The relative error are shown in blue colour.

d) MoDWT-XGB



e) MoDWT-RF



f) MoDWT-MLP

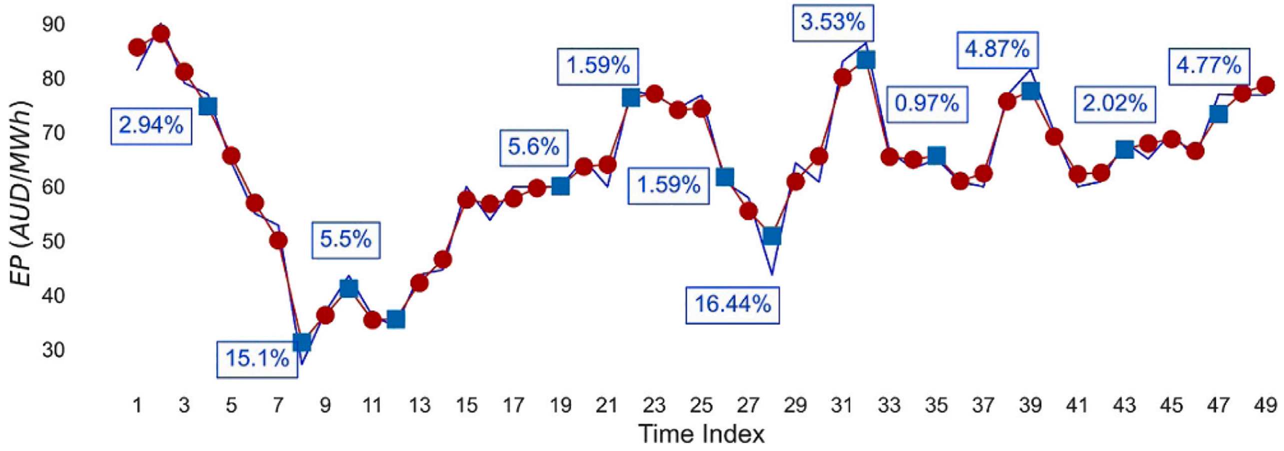


Fig. B.19. (*continued*).

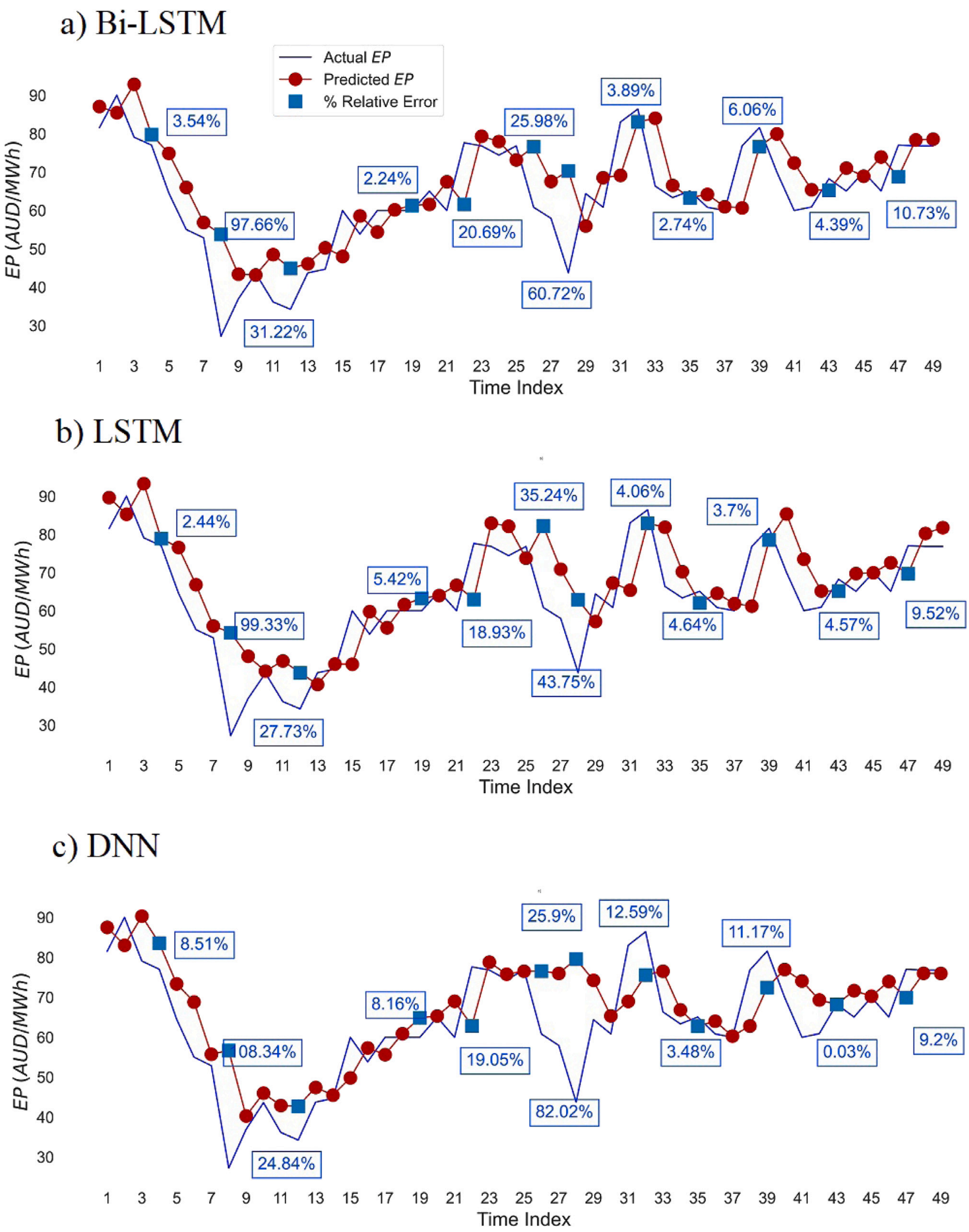
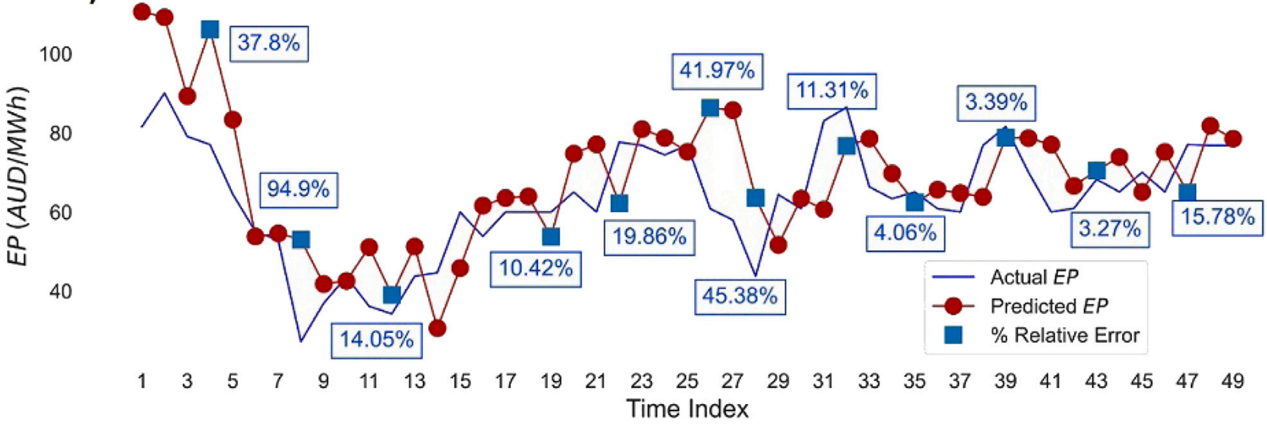
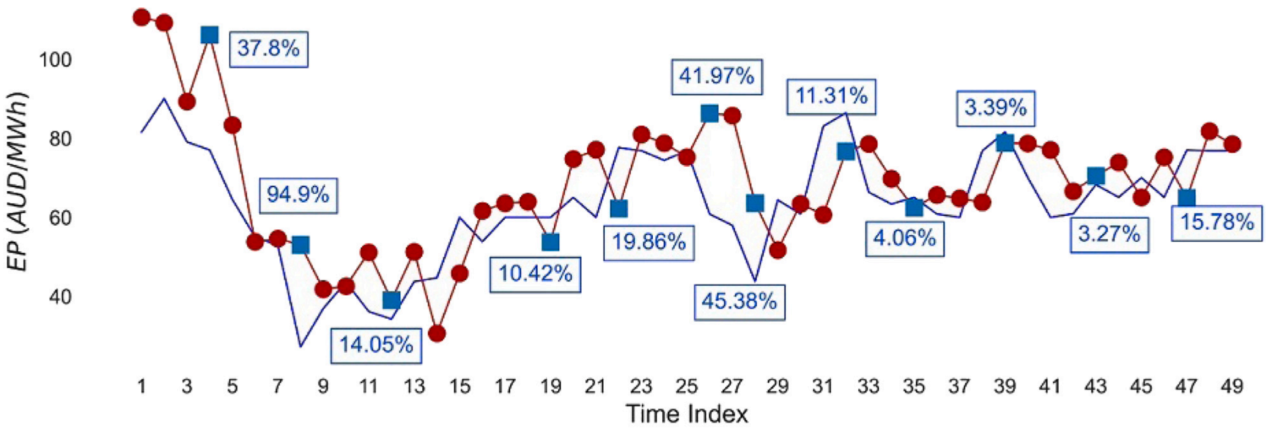


Fig. B.20. Comparative Evaluation of Predicted *vs.* Actual Electricity Prices (AUD/MWh) for the $DS1$ Dataset During a 2-Day Test Period Using a standalone Models. The relative error are shown in blue colour.

d) XGB



e) RF



f) MLP

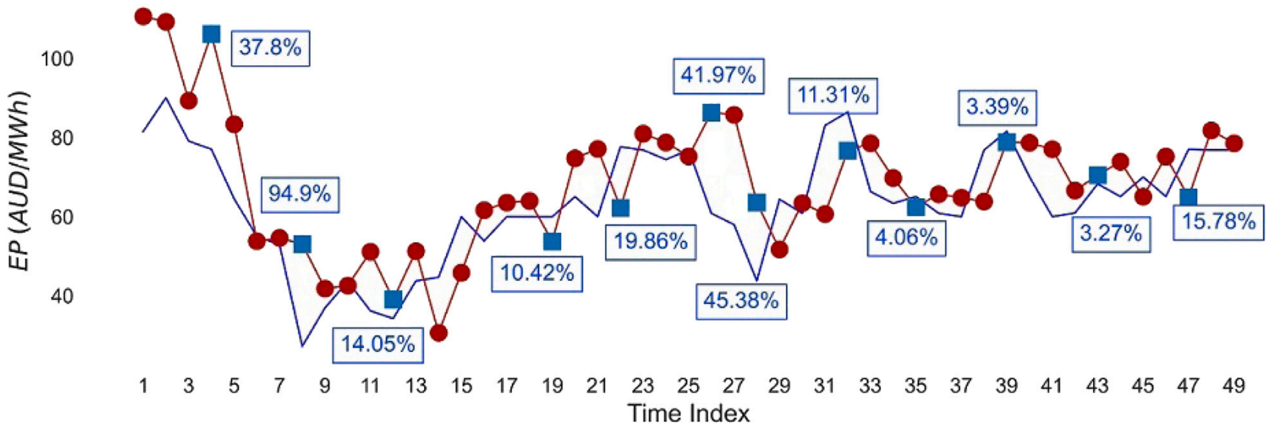


Fig. B.20. (*continued*).

Table A.15
List of Acronyms.

Acronyms	Term
ANFIS	Adaptive Neuro-Fuzzy Inference System
AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Autoregressive
AEMO	Australian Energy Market Operator
ANEM	Australian National Electricity Market
AUD	Australian Dollar
NARX	Auto Regressive Model With Exogenous
AR	Autoregressive
ARIFMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ARXTV	Autoregressive Time Varying
ARNN	Autoregressive Neural Network
ARMA	Autoregressive-Moving-Average
BDL	Bayesian Deep Learning
BO	Bayesian Optimization
BNN	Bayesian Neural Network
BRNN	Bayesian Recurrent Neural Network
BiLSTM	Bi-Directional Long Short-Term Memory
BiLSTM	Bii-Directional Lstm
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CEEMDAN	Complementary Ensemble Empirical Mode Decomposition With Adaptive Noises
CEEMD	Complete Ensemble Empirical Mode Decomposition
CNN	Convolutional Neural Network
DBN	Deep Beleif Network
DL	Deep Learning
DNN	Deep Neural Networks
DFNN	Deep Feedforward Neural Network
DWD	Discrete Wavelet Decomposition
EP	Electricity Prices (Aud/Mwh)
ENN	Elman Neural Network
ENN	Elman Neural Networks
ERNN	Elman Recurrent Neural Network
EMD	Empirical Mode Decomposition
EWT	Empirical Wavelet Transform
EGT	Enhanced Game Theoretic Clustering
ECNN	Enhanced Convolutional Neural Network
ESVM	Enhanced Support Vector Machine
EEMD	Ensemble Empirical Mode Decomposition
EAGNM	Environmentally Adapted Generalized Neuron Model
EWMC	Equal-Weighted Mean Combination
ERC	Error Compensation
ERC-DNN	Error Compensation Deep Neural Network
ES	Exponential Smoothing
EELM	Extended Extreme Learning Machine
XGB	Extreme Gradient Boost
ELM	Extreme Learning Machine
ESMD	Extreme-Point Symmetric Mode Decomposition
FEEMD	Fast Ensemble Empirical Mode Decomposition
FMRVR	Fast Multi-Output Relevance Vector Regression
FFNN	Feed Forward Neural Networks
FR	Fuzzy Regression
GRU	Gated Recurrent Unit
GPR	Gaussian Process Regression
GRNN	General Regression Neural Network
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GRNN	Generalized Regression Neural Network
GRBFN	Generalized Radial Basis Function Neural Network
GA	Genetic Algorithm
GBM	Gradient Boosting Machine
ICEEMDAN	Improved Complementary Ensemble Empirical Mode Decomposition With Adaptive Noises (Improved)
IEMD	Improved Empirical Mode Decomposition
ILRCNN	Integrated Long-Term Recurrent Convolutional Network
KNNR	K-Nearest Neighbors Regression
LASSO	Least Absolute Shrinkage And Selection Operator
LSBoost	Least Squares Regression Boosting
LSSVM	k Least Squares Support Vector Machine
LSTM	Long Short-Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
NBEATS	Neural Basis Expansion Analysis For Interpretable Time Series
NN	Neural Network
NSW	New South Wales
NARMAX	Nonlinear Autoregressive Moving Average Model With Exogenous Inputs
NARNN	Nonlinear Autoregressive Neural Network
OLS	Ordinary Least Squares Regression
PACF	Partial Autocorrelation Function
PSO	Particle Swarm Optimization
RBFNN	Radial Basis Function Neural Network
RF	Random Forest Regression
RVFL	Random Vector Functional Link
RNN	Recurrent Neural Network
RM	Regression Model
RELM	Regularized Extreme Learning Machine
RMSE	Root Mean Square Error
SEQ	South-East Queensland
SDAE	Stacked Denoising Autoencoders
sMAPE	Symmetric Mean Absolute Percentage Error
VHGP	Variational Heteroscedastic Gaussian Process
VMD	Variational Mode Decomposition
VMF	Variational Mode Function
VAR	Vector Autoregressive Model
WNN	Wavelet Neural Network
WPD	Wavelet Packet Denoise
WT	Wavelet Transform
MoDWT	Maximal Overlap Discrete Wavelet Transform

Table C.16
Initial hyperparameters of predictive models are shown for five dataset: $DS1$ (Winter); $DS2$ (Autumn); $DS3$ (Spring); $DS4$ (Summer); and $DS5$ (Yearly 2022).

Predictive models	Model hyperparameters	*DS1*	*DS2*	*DS3*	*DS4*	*DS5*
Convolution Neural Network Integrated with Random Vector Functional Link (CRVFL)	Filter 1 (CNN)	***70***	***55***	***45***	***45***	***40***
	Filter 2 (CNN)	***85***	***95***	***75***	***65***	***65***
	Filter 3 (CNN)	***70***	***40***	***40***	***35***	***25***
	Filter 4 (CNN)	***20***	***25***	***35***	***25***	***20***
	Epochs (CNN)	***1000***	***1000***	***1000***	***1000***	***1000***
	Activation function	***ReLU***	***ReLU***	***ReLU***	***ReLU***	***ReLU***
	Solver	***Adam***	***Adam***	***Adam***	***Adam***	***Adam***
	Batch Size	***250***	***850***	***650***	***1050***	***850***
	Enhancement nodes	***27***	***9***	***19***	***13***	***37***
	Regularization parameter	***1.0***	***1.5***	***0.5***	***1.5***	***1.0***
Deep Neural Network (DNN)	Hiddenneuron 1	***80***	***80***	***50***	***80***	***40***
	Hiddenneuron 2	***75***	***65***	***50***	***60***	***60***
	Hiddenneuron 3	***50***	***55***	***50***	***65***	***75***
	Batch Size	***300***	***300***	***1050***	***650***	***850***
	Solver	***Adam***	***Adam***	***Adam***	***Adam***	***Adam***
	Epochs	***1000***	***1000***	***1000***	***1000***	***1000***
Random Forest Regression (RF)	The maximum depth of the tree.	***19***	***18***	***18***	***14***	***12***
	The number of trees in the forest.	***77***	***63***	***67***	***81***	***63***
	Minimum number of samples to split an internal node	***5***	***5***	***4***	***3***	***12***
	The number of features to consider when looking for the best split.	***auto***	***auto***	***auto***	***auto***	***auto***
Long Short Term Memory Network (LSTM) and Bidirectional LSTM (Bi-LSTM)	LSTM cell 1	***55***	***55***	***75***	***95***	***65***
	LSTM cell 2	***60***	***75***	***75***	***65***	***80***
	Activation function	***ReLU***	***ReLU***	***ReLU***	***ReLU***	***ReLU***
	Epochs	***1000***	***1000***	***1000***	***1000***	***1000***
	Drop rate	***0***	***0***	***0.5***	***0***	***0.5***
	Batch Size	***1050***	***1150***	***750***	***850***	***750***
eXtreme Gradient Boosting (XGB)	Booster Type	***gbtree’***	***gbtree’***	***gbtree’***	***gbtree’***	***gbtree’***
	Step size shrinkage used in update to prevents overfitting.	***0.3***	***0.3***	***0.1***	***0.5***	***0.5***
	The maximum depth of the tree.	***8***	***12***	***9***	***14***	***9***
	The number of trees in the forest.	***87***	***69***	***57***	***95***	***69***
Multi Layer Perceptron (MLP)	Hidden neuron	***60***	***60***	***80***	***90***	***80***
	Activation function	***ReLU***	***ReLU***	***ReLU***	***ReLU***	***ReLU***
	Learning rate	***0.001***	***0.001***	***0.001***	***0.006***	***0.001***
	Solver	***Adam***	***Adam***	***Adam***	***Adam***	***Adam***

Note: ‘gbtree’ = Gradient Boosted Trees, ReLU = Rectified Linear Unit Activation Function, and Adam = Adaptive Moment Estimation.

References

[1] Weron R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. Int J Forecast 2014;30(4):1030–81.

[2] 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.

[3] Reddy SS, Momoh JA. Short term electrical load forecasting using back propagation neural networks. In: 2014 North American power symposium. IEEE; 2014, p. 1–6.

[4] Marcjasz G, Uniejewski B, Weron R. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. Int J Forecast 2019;35(4):1520–32.

[5] Kiose D, Voudouris V. The ACEWEM framework: An integrated agent-based and statistical modelling laboratory for repeated power auctions. Expert Syst Appl 2015;42(5):2731–48.

[6] Burger M, Graeber B, Schindlmayr G. Managing energy risk: An integrated view on power and other energy markets. 2014.

[7] Islyaev S, Date P. Electricity futures price models: Calibration and forecasting. European J Oper Res 2015;247(1):144–54.

[8] González JP, San Roque AMSM, Perez EA. Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. IEEE Trans Power Syst 2017;33(1):545–56.

[9] Peter SE, Raglend IJ. Sequential wavelet-ANN with embedded ANN-pso hybrid electricity price forecasting model for Indian energy exchange. Neural Comput Appl 2017;28:2277–92.

[10] Memarzadeh G, Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electr Power Syst Res 2021;192:106995.

[11] Gurtler M, Paulsen T. Forecasting performance of time series models on electricity spot markets. Int J Energy Sector Manag 2018;12(4):617–40.

[12] 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.

[13] Ioannidis F, Kosmidou K, Savva C, Theodossiou P. Electricity pricing using a periodic GARCH model with conditional skewness and kurtosis components. Energy Econ 2021;95:105110.

[14] Lago J, De Ridder F, Vrancx P, De Schutter B. Forecasting day-ahead electricity prices in Europe: The importance of considering market integration. Appl Energy 2018;211:890–903.

[15] Szkuta BR, Sanabria LA, Dillon TS. Electricity price short-term forecasting using artificial neural networks. IEEE Trans Power Syst 1999;14(3):851–7.

[16] Lin W-M, Gow H-J, Tsai M-T. An enhanced radial basis function network for short-term electricity price forecasting. Appl Energy 2010;87(10):3226–34.

[17] Yamin H, Shahidehpour S, Li Z. Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets. Int J Electr Power Energy Syst 2004;26(8):571–81.

[18] Ghayekhloo M, Azimi R, Ghofrani M, Menhaj M, Shekari E. A combination approach based on a novel data clustering method and Bayesian recurrent neural network for day-ahead price forecasting of electricity markets. Electr Power Syst Res 2019;168:184–99.

[19] Che J, Wang J. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. Energy Convers Manage 2010;51(10):1911–7.

[20] Ludwig N, Feuerriegel S, Neumann D. Putting big data analytics to work: Feature selection for forecasting electricity prices using the LASSO and random forests. J Decis Syst 2015;24(1):19–36.

[21] Chaâbane N. A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. J Appl Stat 2014;41(3):635–51.

[22] Catalão JPdS, Pousinho HMI, Mendes VMF. Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. IEEE Trans Power Syst 2010;26(1):137–44.

[23] Pindoriya N, Singh S, Singh S. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans Power Syst 2008;23(3):1423–32.

[24] Zhang J, Tan Z, Li C. A novel hybrid forecasting method using GRNN combined with wavelet transform and a GARCH model. Energy Sources, Part B: Econ, Plan Policy 2015;10(4):418–26.

[25] Chen X, Dong ZY, Meng K, Xu Y, Wong KP, Ngan H. Electricity price forecasting with extreme learning machine and bootstrapping. IEEE Trans Power Syst 2012;27(4):2055–62.

[26] Keles D, Scelle J, Paraschiv F, Fichtner W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl Energy 2016;162:218–30.

[27] Rafiei M, Niknam T, Khooban M-H. Probabilistic forecasting of hourly electricity price by generalization of ELM for usage in improved wavelet neural network. IEEE Trans Ind Inf 2016;13(1):71–9.

[28] Wang L, Tao R, Hu H, Zeng Y-R. Effective wind power prediction using novel deep learning network: Stacked independently recurrent autoencoder. Renew Energy 2021;164:642–55.

[29] Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling. In: Thirteenth annual conference of the international speech communication association. 2012.

[30] Brusaferri A, Matteucci M, Portolani P, Vitali A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Appl Energy 2019;250:1158–75.

[31] Dedinec A, Filiposka S, Dedinec A, Kocarev L. Deep belief network based electricity load forecasting: An analysis of Macedonian case. Energy 2016;115:1688–700.

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

[33] Yang H, Schell KR. Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. Appl Energy 2021;299:117242.

[34] Li W, Becker DM. Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. Energy 2021;237:121543.

[35] Deng Z, Wang B, Xu Y, Xu T, Liu C, Zhu Z. Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting. IEEE Access 2019;7:88058–71.

[36] Chinnathambi RA, Plathottam SJ, Hossen T, Nair AS, Ranganathan P. Deep neural networks (DNN) for day-ahead electricity price markets. In: 2018 IEEE electrical power and energy conference. IEEE; 2018, p. 1–6.

[37] Jiang L, Hu G. Day-ahead price forecasting for electricity market using long-short term memory recurrent neural network. In: 2018 15th international conference on control, automation, robotics and vision. IEEE; 2018, p. 949–54.

[38] Zhu Y, Dai R, Liu G, Wang Z, Lu S. Power market price forecasting via deep learning. In: IECON 2018-44th annual conference of the IEEE industrial electronics society. IEEE; 2018, p. 4935–9.

[39] Mujeeb S, Javaid N, Akbar M, Khalid R, Nazeer O, Khan M. Big data analytics for price and load forecasting in smart grids. In: Advances on broadband and wireless computing, communication and applications: proceedings of the 13th international conference on broadband and wireless computing, communication and applications. Springer; 2019, p. 77–87.

[40] Zahid M, Ahmed F, Javaid N, Abbasi RA, Zainab Kazmi HS, Javaid A, et al. Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids. Electronics 2019;8(2):122.

[41] Wang D, Luo H, Grunder O, Lin Y, Guo H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl Energy 2017;190:390–407.

[42] Fan S, Mao C, Chen L. Next-day electricity-price forecasting using a hybrid network. IET Gener, Transm Distrib 2007;1(1):176–82.

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

[44] Lago J, Marcjasz G, De Schutter B, Weron R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Appl Energy 2021;293:116983.

[45] Xiao L, Shao W, Yu M, Ma J, Jin C. Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting. Appl Energy 2017;198:203–22.

[46] 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.

[47] Itaba S, Mori H. A fuzzy-preconditioned GRBFN model for electricity price forecasting. Procedia Comput Sci 2017;114:441–8.

[48] Zhang J-L, Zhang Y-J, Li D-Z, Tan Z-F, Ji J-F. Forecasting day-ahead electricity prices using a new integrated model. Int J Electr Power Energy Syst 2019;105:541–8.

[49] Heydari A, Nezhad MM, Pirshayan E, Garcia DA, Keynia F, De Santoli L. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. Appl Energy 2020;277:115503.

[50] Naz A, Javed MU, Javaid N, Saba T, Alhussein M, Aurangzeb K. Short-term electric load and price forecasting using enhanced extreme learning machine optimization in smart grids. Energies 2019;12(5):866.

[51] 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.

[52] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl Energy 2017;190:291–305.

[53] Ghofrani M, Azimi R, Najafabadi F, Myers N. A new day-ahead hourly electricity price forecasting framework. In: 2017 North American power symposium. IEEE; 2017, p. 1–6.

[54] Gobu B, Jaikumar S, Arulmozhi N, Kanimozhi P, et al. Two-stage machine learning framework for simultaneous forecasting of price-load in the smart grid. In: 2018 17th IEEE international conference on machine learning and applications. IEEE; 2018, p. 1081–6.

[55] Gao W, Darvishan A, Toghani M, Mohammadi M, Abedinia O, Ghadimi N. Different states of multi-block based forecast engine for price and load prediction. Int J Electr Power Energy Syst 2019;104:423–35.

[56] Zhou L, Wang B, Wang Z, Wang F, Yang M. Seasonal classification and RBF adaptive weight based parallel combined method for day-ahead electricity price forecasting. In: 2018 IEEE power & energy society innovative smart grid technologies conference. IEEE; 2018, p. 1–5.

[57] Osório GJ, Lotfi M, Shafie-Khah M, Campos VM, Catalão JP. Hybrid forecasting model for short-term electricity market prices with renewable integration. Sustainability 2018;11(1):57.

[58] Abu-Elanien AE, Abdel-Khalik AS, Massoud AM, Ahmed S. A non-communication based protection algorithm for multi-terminal HVDC grids. Electr Power Syst Res 2017;144:41–51.

[59] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manage 2018;159:54–64.

[60] Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: 2011 IEEE international conference on acoustics, speech and signal processing. IEEE; 2011, p. 4144–7.

[61] Ghimire S, Deo RC, Raj N, Mi J. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. Renew Sustain Energy Rev 2019;113:109247.

[62] Ghimire S, Deo RC, Casillas-Perez D, Salcedo-Sanz S. Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. Renew Energy 2022;190:408–24.

[63] Rani R, Jessie H, Victoire T, Albert A. A hybrid elman recurrent neural network, group search optimization, and refined VMD-based framework for multi-step ahead electricity price forecasting. Soft Comput 2019;23(18):8413–34.

[64] Cornish CR, Bretherton CS, Percival DB. Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. Bound-Layer Meteorol 2006;119:339–74.

[65] Nury AH, Hasan K, Alam MJB. Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh. J King Saud Univ-Sci 2017;29(1):47–61.

[66] Deléchelle E, Lemoine J, Niang O. Empirical mode decomposition: An analytical approach for sifting process. IEEE Signal Process Lett 2005;12(11):764–7.

[67] Zhao Y, Cui H, Huo H, Nie Y. Application of synchrosqueezed wavelet transforms for extraction of the oscillatory parameters of subsynchronous oscillation in power systems. Energies 2018;11(6):1525.

[68] Panja M, Chakraborty T, Nadim SS, Ghosh I, Kumar U, Liu N. An ensemble neural network approach to forecast dengue outbreak based on climatic condition. Chaos Solitons Fractals 2023;167:113124.

[69] Baulon L, Allier D, Massei N, Bessiere H, Fournier M, Bault V. Influence of low-frequency variability on groundwater level trends. J Hydrol 2022;606:127436.

[70] Massei N, Dieppois B, Hannah D, Lavers D, Fossa M, Laignel B, et al. Multi-time-scale hydroclimate dynamics of a regional watershed and links to large-scale atmospheric circulation: Application to the seine river catchment, France. J Hydrol 2017;546:262–75.

[71] Percival DB, Walden AT. Wavelet methods for time series analysis, vol. 4, Cambridge University Press; 2000.

[72] Ghimire S, Nguyen-Huy T, Prasad R, Deo RC, Casillas-Pérez D, Salcedo-Sanz S, et al. Hybrid convolutional neural network-multilayer perceptron model for solar radiation prediction. Cogn Comput 2022;1–27.

[73] 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;1–38.

[74] Wang F, Zhang Z, Liu C, Yu Y, Pang S, Duić N, et al. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. Energy Convers Manag 2019;181:443–62.

[75] Ghimire S, Bhandari B, Casillas-Pérez D, Deo RC, Salcedo-Sanz S. Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in queensland, Australia. Eng Appl Artif Intell 2022;112:104860.

[76] Ghimire S, Nguyen-Huy T, AL-Musaylh MS, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. A novel approach based on integration of convolutional neural networks and echo state network for daily electricity demand prediction. Energy 2023;127430.

[77] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 2017;10(1):841–51.

[78] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognit 2018;77:354–77.

[79] Zou Y, Yu L, Tso GK, He K. Risk forecasting in the crude oil market: A multiscale convolutional neural network approach. Phys A 2020;541:123360.

[80] Pao Y-H, Takefuji Y. Functional-link net computing: Theory, system architecture, and functionalities. Computer 1992;25(5):76–9.

[81] Pao Y-H, Park G-H, Sobajic DJ. Learning and generalization characteristics of the random vector functional-link net. Neurocomputing 1994;6(2):163–80.

[82] Zhang L, Suganthan PN. A comprehensive evaluation of random vector functional link networks. Inf Sci 2016;367:1094–105.

[83] Ren Y, Suganthan PN, Srikanth N, Amaratunga G. Random vector functional link network for short-term electricity load demand forecasting. Inform Sci 2016;367:1078–93.

[84] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35(8):1798–828.

[85] Ghimire S, Yaseen ZM, Farooque AA, Deo RC, Zhang J, Tao X. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep 2021;11(1):1–26.

[86] Ghimire S, Deo RC, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 2019;253:113541.

[87] Ghimire S, Nguyen-Huy T, Deo RC, Casillas-Perez D, Salcedo-Sanz S. Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. Sustain Mater Technol 2022;32:e00429.

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

[89] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016, p. 785–94.

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

[91] Ghimire S, Deo RC, Wang H, Al-Musaylh MS, Casillas-Pérez D, Salcedo-Sanz S. Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results. Energies 2022;15(3):1061.

[92] Jarque CM, Bera AK. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Econ Lett 1980;6(3):255–9.

[93] Cheung Y-W, Lai KS. Lag order and critical values of the augmented Dickey–Fuller test. J Bus Econom Statist 1995;13(3):277–80.

[94] Brock WA, Dechert WD, Scheinkman JA, LeBaron B. A test for independence based on the correlation integral. Technical report, University of Wisconsin-Madison; 1986.

[95] Juhani JM, Ibrahim R. Mother wavelet selection for control valve leakage detection using acoustic emission. In: 2018 IEEE conference on systems, process and control. IEEE; 2018, p. 224–7.

[96] Kankar PK, Sharma SC, Harsha SP. Fault diagnosis of ball bearings using continuous wavelet transform. Appl Soft Comput 2011;11(2):2300–12.

[97] Kankar PK, Sharma SC, Harsha SP. Rolling element bearing fault diagnosis using wavelet transform. Neurocomputing 2011;74(10):1638–45.

[98] de Oliveira HM. Shannon and renyi entropy of wavelets. 2015, arXiv preprint arXiv:1502.01871.

[99] Khalighi S, Sousa T, Pires G, Nunes U. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels. Expert Syst Appl 2013;40(17):7046–59.

[100] Granata F, Di Nunno F, de Marinis G. Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: A comparative study. J Hydrol 2022;613:128431.

[101] 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.

[102] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172(5):1122–31.

[103] 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;100302.

[104] 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.

[105] 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.

[106] Al-Musaylh MS, Deo RC, Adamowski JF, Li Y. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF reanalysis atmospheric predictors in southeast Queensland, Australia. Renew Sustain Energy Rev 2019;113:109293.