A Two-Stage Deep Learning Approach for Accurate Day-Ahead Electricity Price Forecasting

Aksu, Inayet Özge; Ghaemi, Sina; Anvari-Moghaddam, Amjad

Published in:
Engineering Applications of Artificial Intelligence

DOI (link to publication from Publisher):
10.1016/j.engappai.2025.112721

Publication date:
2026

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Aksu, I. Ö., Ghaemi, S., & Anvari-Moghaddam, A. (2026). A Two-Stage Deep Learning Approach for Accurate Day-Ahead Electricity Price Forecasting. *Engineering Applications of Artificial Intelligence*, *163*, Article 112721. https://doi.org/10.1016/j.engappai.2025.112721

Engineering Applications of Artificial Intelligence 163 (2026) 112721

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A two-stage deep learning approach for accurate day-ahead electricity price forecasting

İnayet Özge Aksu [a,b], Sina Ghaemi [b], Amjad Anvari-Moghaddam [b,*]

[a] Department of Artificial Intelligence Engineering, Adana Alparslan Türkeş Science and Technology University, 01250, Adana, Türkiye
[b] Department of Energy (AAU Energy), Aalborg University, 9220, Aalborg, Denmark

ARTICLE INFO

Keywords:
Two-stage prediction
Deep learning
Hybrid
Decomposition
Day-ahead
Energy market

ABSTRACT

Participants in the energy market are at greater risk of making decisions due to the nonlinear and volatile characteristics of electricity prices. Accurate short-term electricity price forecasting (EPF) is essential to ensure improved resource allocation, grid stability and enable market participants to manage their decisions efficiently. This study proposes a novel two-stage forecasting framework for day-ahead EPF using time series decomposition methods and hybrid deep learning algorithms. In the first stage, features related to EPF at the next time step are predicted. In this stage, the highest-frequency component extracted via Empirical Mode Decomposition (EMD) is further decomposed using Variational Mode Decomposition (VMD) so as to better capture rapid fluctuations and improve the overall prediction accuracy. Moreover, a decentralized deep learning architecture is designed in which Gated Recurrent Unit (GRU) networks are employed for high-frequency components, while Long Short-term Memory (LSTM) networks are used for the remaining components. In the second stage, EPF is generated using a hybrid LSTM and GRU structure, which incorporates both features estimated in the first stage and historical electricity price data. Finally, hyperparameters of the deep learning models are optimized using Bayesian Optimization to enhance performance. To validate the proposed framework, real market data from the DK1 region of Denmark is used. The proposed hybrid prediction framework is evaluated against both machine learning methods and deep learning-based architectures. Experimental results demonstrate that the proposed method achieves approximately 27.15 % lower RMSE compared to traditional machine learning models, and around 28.24 % lower RMSE compared to LSTM-based models.

1. Introduction

1.1. Motivation

In the competitive electricity market, accurate electricity price forecasting (EPF) is critically important for market participants engaged in buying and selling transactions in order to maximize their profit. Market participants rely on price forecasts to determine their bidding strategies, allocate resources efficiently, and plan facility investments (Gong et al., 2025). In such a competitive environment, any participant who can accurately predict future electricity prices can gain additional advantage and achieves greater profits. The supply-demand relationship among participants plays a key role in determining electricity prices under market conditions. These decisions are heavily influenced by the participants' own price forecasts. Therefore, the ability to act efficiently in the market depends on the accurate implementation of EPF.

1.2. Literature review

Recent studies show that deep learning, machine learning, and intelligent algorithms are increasingly favored for electricity price prediction. Researchers primarily aim to achieve accurate forecasts despite the presence of price volatility and sudden spikes. Earlier studies in this field largely relied on statistical models particularly time series approaches such as Autoregressive Integrated Moving Average (ARIMA) (Rajan and Chandrakala, 2021), (McHugh et al., 2022), (Abdellatif et al., 2023) and Generalized Autoregressive Conditional Heteroskedastic (GARCH) (da Silva Leite and de Lima, 2023), (Janczura and Puć, 2023). However, these methods often fall short in capturing the complex and non-linear dynamics of electricity markets, making them less responsive to abrupt price changes. ARIMA methods are successful in modeling linear relationships and seasonal trends and are suitable for systems that do not show major changes over time and operate in a predictable and

* Corresponding author. Aalborg University, Pontoppidanstraede 111, room 1.164, 9220, Aalborg, Denmark.
E-mail addresses: oaksu@atu.edu.tr, ioa@energy.aau.dk (İ.Ö. Aksu), sigh@energy.aau.dk (S. Ghaemi), aam@energy.aau.dk (A. Anvari-Moghaddam).

https://doi.org/10.1016/j.engappai.2025.112721
Received 22 May 2025; Received in revised form 6 September 2025; Accepted 8 October 2025
Available online 9 November 2025

balanced manner. However, they are inadequate in modeling the complex and variable structure in energy systems. In order to increase the prediction accuracy in future studies, Computational Intelligence (CI) models and hybrid models have begun to be used. Initially, in the CI field, intelligent system methods were frequently preferred for EPF. In intelligence systems, Artificial Neural Network (ANN) (Panapakidis and Dagoumas, 2016), Fuzzy (Plakas et al., 2023) and Neuro-fuzzy (Setayesh Nazar and Eslami Fard, 2021) and, Random Forest (de Castilho Braz et al., 2024) methods have been used in electricity price forecasting studies. These methods can capture nonlinear dependencies and integrate external factors into the models used in the estimation phase. The use of these methods in the EPF field has increased over time.

Later, with the use of big data in the EPF field, these models, despite their strengths, were insufficient to extract complex relationships in the data. Models such as ARIMA and simple ANNs can generally only see surface relationships; They are insufficient to capture deeper relationships in the data related to electricity prices. For this reason, deep learning methods have begun to be used in EPF studies. The preferred deep learning methods in EPF can be Deep Neural Network (DNN) (Huang et al., 2021), Support Vector Machine (SVM) (Zhang et al., 2020), Recurrent Neural Network (RNN) (Kaya et al., 2023), Gated Recurrent Unit (GRU) (Yang and Schell, 2020), Long Short-term Memory (LSTM) (Wang et al., 2023), and Convolution Neural Network (CNN) (Khan et al., 2020).

Previously some of the stand-alone models used in the EPF field were successful under certain conditions. However, it became clear that these methods were unable to fully capture the deep relationships in electricity price data with complex, volatile, and non-stationary structure (i.e., statistically changing over time). To overcome this difficulty, researchers began to develop "hybrid" models that combine multiple methods. A hybrid model is one that combines several different methods, and as a result, it becomes more powerful and flexible than a single model, and is therefore capable of producing more accurate and reliable results. Moradzadeh et al. in (Moradzadeh et al., 2025) aimed at short-term electricity price forecasting using hourly Ontario energy price data from Canada and they proposed a hybrid deep learning model. In this paper, Bidirectional LSTM and Gated Recurrent Unit are hybridized to improve the prediction performance. The obtained results were compared with Extreme Learning Machine (ELM), CNN-LSTM and Autoencoder + BiLSTM (AE-BiLSTM) to demonstrate the success of the proposed method. Similarly, in (Pourdaryaei et al., 2024), a method that can model both spatial and temporal relationships in market fluctuations more effectively by using multi-head attention and 1D-CNN together with feature selection for short-term electricity price prediction was developed. With the proposed model, low error rates were obtained in the analysis for each season.

The use of hybrid models provided high estimates in complex and dynamic EPF. Due to the complex structure of hybrid models and the fact that parameter tuning involves a large number of possibilities, the use of methods such as Bayesian optimization in parameter tuning increases the predictive performance of the model. In (Dai and Yu, 2024), the parameters of prediction model developed by integrating CNN, TCN, Attention mechanism were optimized by Bayesian optimization method. Jayanth and Manimaran in (Jayanth and Manimaran, 2024) have optimized the parameters of the proposed hybrid forecasting method (Double Exponential Smoothing + Dual Attention Encoder-Decoder + Bi-Directional GRU) in order to maximize the performance of the model using Bayesian Optimization.

These methods assist only in tuning parameters that lie outside the learning process-that is, parameters the model cannot learn on its own and must be predefined. However, tuning these external parameters alone is not sufficient to ensure the model's efficiency. To address these limitations and enhance model performance, time series decomposition techniques have been integrated with deep learning algorithms. Another hybridization option to deal with the complexity of EPF time series data is the use of time series decomposition methods such as Wavelet Transform (WF) (Osório et al., 2018), Empirical Mode Decomposition (EMD) (Zhang et al., 2021), Variational Mode Decomposition (VMD) (Xiong and Qing, 2023), (Xu et al., 2025), Ensemble EMD (EEMD) (Khan et al., 2021), etc. In this regard, Xiong and Qing (2023) have proposed a new hybrid method that, after the feature selection step, performs signal decomposition on the selected features using the VMD method for noise reduction and information extraction. The parameters of the LSTM model used in the estimation method were optimized with Bayesian Optimization and Hyperband. Likewise, in (Tan et al., 2023), time series data is decomposed into low-frequency and high-frequency components with ICEEMDAN to better model the complex volatility of prices in day-ahead electricity price forecasting for the Australian national electricity market (in NSW). In addition, Inspired Grey Wolf Optimizer (IGWO) was used for hyperparameter optimization in this study.

1.3. Contribution of this work

This research proposes a two-stage day-ahead electricity price forecasting (EPF) algorithm, recognizing the importance of EPF for both researchers and industry stakeholders. In the first stage, the input features for price prediction are decomposed into frequency components using EMD. The most complex component is further processed using VMD and modeled with GRU, while the remaining components are modeled with a LSTM network. The results are then combined to generate a precise t+1 price prediction in the next stage. Indeed, in the second stage, a 24-h electricity price forecast is performed using the t+1 feature values predicted in the first stage, applying a decentralized LSTM + GRU framework. Finally, the main contributions of the paper can be summarized as follows:

- A novel two-stage hybrid forecasting framework is introduced, combining signal decomposition with a heterogeneous deep learning architecture to separately model low- and high-frequency components of electricity price time series, enabling component-specific learning and improved forecast accuracy.
- A decentralized deep learning structure is developed, integrating LSTM and GRU networks to simultaneously capture long-term dependencies and short-term fluctuations in price signals, optimized through Bayesian methods for enhanced predictive performance.
- Comprehensive real-world validation is conducted using DK1 market data from Denmark, demonstrating the proposed model's superiority over conventional ML/DL methods in terms of accuracy and robustness in day-ahead electricity price forecasting.

The rest of the paper is structured as follows. Section 2 introduces the proposed framework and presents details regarding the dataset, performance metrics, and hyperparameter optimization process. Section 3 provides a comprehensive analysis of each stage within the two-stage estimation process and compares the results with benchmark methods using various performance metrics. Finally, Section 4 summarizes the key findings of the study and offers recommendations for future research directions.

2. Method of research

The effectiveness of accurate price forecasting can be assessed from the perspectives of both grid producers and consumers. For producers, precise forecasts support informed production allocation decisions, enabling optimal use of generation resources across regions to maximize profitability. On the other hand, consumers depend on accurate price predictions to plan budgets, manage consumption patterns, and mitigate the impact of price volatility. In today's increasingly volatile energy markets, the ability to anticipate price fluctuations with high accuracy is essential-not only for securing profit but also for avoiding unexpected financial losses. To address this challenge, this section proposes a new framework that employs a hybrid two-stage forecasting approach

designed to enhance predictive accuracy. The proposed model consists of two main stages: (i) Multi-step feature forecasting (multi-step forecasting of selected features) and (ii) Electricity price prediction model for the specific bidding zone. (day-ahead price prediction using these forecasted inputs). The following sections provide detailed explanations of the framework consisting of these two steps. An overall schematic representation of the proposed methodology is illustrated in Fig. 1.

2.1. Introduction of decomposition techniques used in the developed model

Time series of key variables in energy markets are often complex and consist of multiple frequency components. Decomposition techniques are employed to decompose the time series into sub-series with distinct frequency characteristics. The goal is to break down the time series into simpler components, allowing each to be analyzed separately for better interpretability and predictive accuracy. Decomposition techniques can generally be applied in either a single-step or multi-step approach. In single-step decomposition, a single method is used to decompose the time series into sub-series based on frequency components. However, a single decomposition method may not fully capture all the structural features of the series. Therefore, hybrid decomposition methods are employed to leverage the strengths of multiple techniques. In multi-step decomposition, additional methods are applied after the initial decomposition to extract all relevant structural characteristics of the time series more effectively. In this study, EMD and VMD techniques are preferred.

2.1.1. Empirical Mode Decomposition (EMD)

EMD is a signal decomposition technique widely used in conjunction with deep learning methods to analyze and process nonlinear and non-stationary data (Huang et al., 1998). This method was developed as an alternative to theoretical decomposition techniques, such as Fourier decomposition, to address the challenge of selecting an appropriate decomposition method for non-stationary signals with time-varying mechanisms (Taheri et al., 2021). When applied to complex real-world signals, EMD decomposes the original signal into multiple simple components known as Intrinsic Mode Functions (IMFs) and one residue component. Two conditions are taken into account when decomposing the signal into IMF components: (i) in each IMF component, the number of local extrema (maximum and minimum values) and the number of zero crossing points should be either equal or differ by at most one, (ii) the signal should be symmetric with respect to the local zero mean. The process of the decomposition includes 5 steps which can be explained as follows:

Step 1: The local maximum and minimum points of the original signal ($X(t)$) are determined.

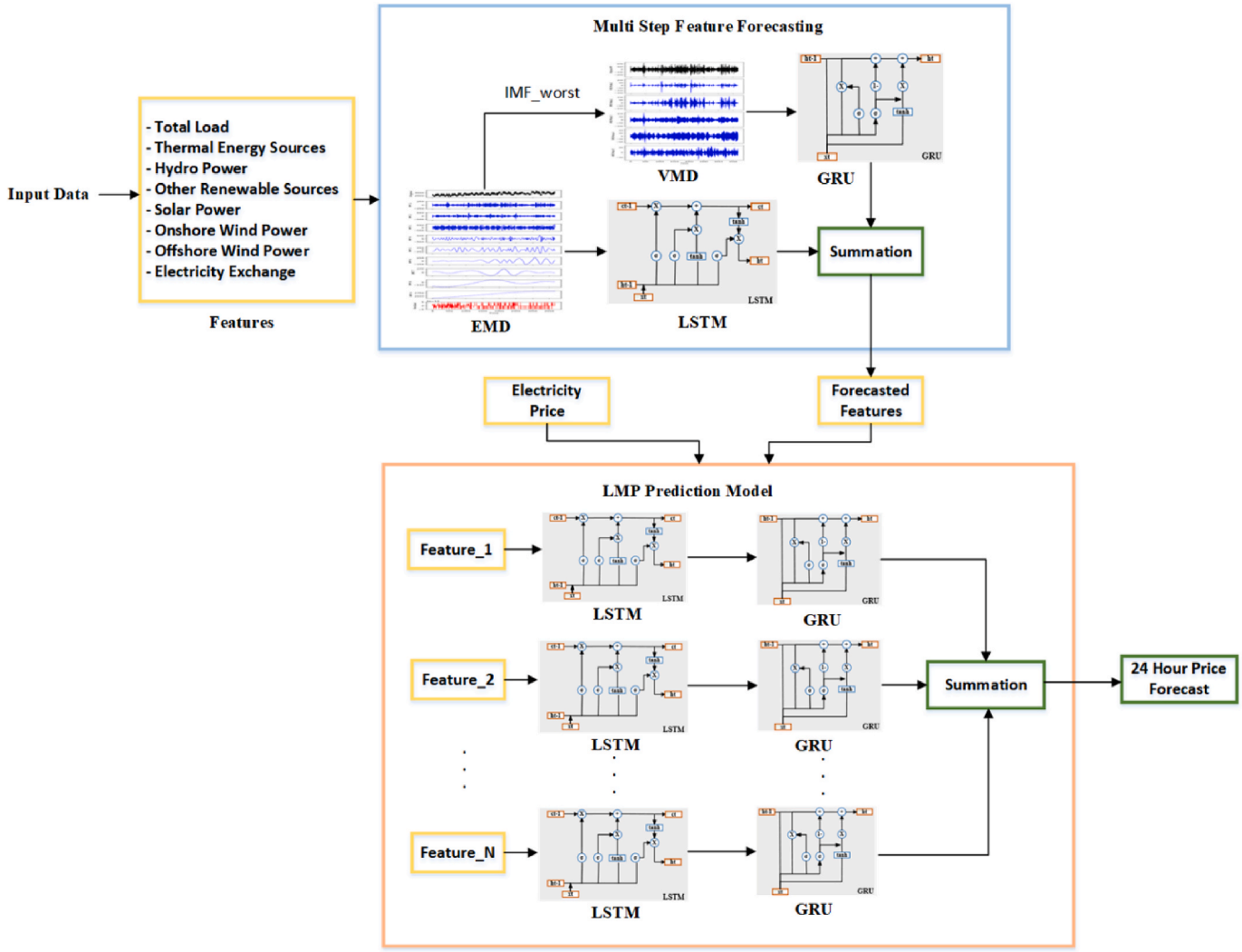

Fig. 1. A framework for forecasting the day-ahead locational marginal price.

Step 2: The local minimum ($e_{min}(t)$) and maximum ($e_{max}(t)$) points required to form the envelope are determined and interpolated between these points.
Step 3: The average envelope is calculated using Eq. (1). The formula for this step is as follows:

$$M(t) = [e_{min}(t) + e_{max}(t)]/2 \tag{1}$$

Step 4: To test whether the new signal is IMF, the mean value is subtracted from the original signal using Eq. (2)

$$S(t) = X(t) - M(t) \tag{2}$$

Step 5: Finally, if the resulting signal satisfies the IMF criteria (zero crossing point and symmetry), it is considered an IMF and set $IMF_1(t) = S(t)$. If not, $S(t)$ is used as the new input signal and we return to the first step. Finally, the mathematical formulation of EMD would be as follows:

$$X(t) = \sum_{n \in N} IMF_n(t) + r_n(t) \tag{3}$$

where, $X(t)$ is the original signal data. $IMF_n(t)$ and $r_n(t)$ denote n^{th} intrinsic mode function (IMF) and the residues, respectively. The residual component is defined as the remaining signal after subtracting the IMFs. Since IMF_1 represents the highest frequency component of the original signal, it contains the most random noise and rapidly varying elements. Therefore, IMF_1 is regarded as the noisiest component.

2.1.2. Variational Mode Decomposition (VMD)

VMD is an adaptive, non-recursive signal processing method (Dragomiretskiy and Zosso, 2013) that decomposes a real-valued signal into a finite number of sub-signals. Each component is nearly compact around its corresponding center frequency. This decomposition technique effectively addresses the limitations of EMD, such as mode mixing and endpoint effects. The signals decomposed using VMD exhibit high accuracy and strong noise robustness. If one seeks to decompose a real-valued signal $X(t)$ into K VMFs each component x_k must be concentrated around its corresponding center frequency w_k. The different steps of this decomposition technique are mentioned as follows:

Step 1: To obtain a single-sided frequency spectrum, the correlation analysis signal for each mode (u_k) is calculated using the Hilbert transform.
Step 2: The frequency spectra obtained for each mode are positioned to the baseband using an exponential function adjusted by the estimated center frequency.
Step 3: In this step, the method transforms the decomposition problem into an optimization problem to determine the modes using Eq. (4):

$$\begin{aligned} &min_{w_k,u_k} \sum_k \left\| \varphi_t \left[\left(\mu(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \\ &s.t. \sum_k u_k = X(t) \end{aligned} \tag{4}$$

where $X(t)$ is the original time series signal. μ represents the Dirac distribution and * represents the convolution operation. u_k and w_k denote the modes and the center pulsations, respectively. $u_k(t)$ shows the mode functions.

2.2. Introduction of deep learning networks

Deep learning methods are mainly used in nonlinear and time-varying time series activities. The basic method of this field is Recurrent Neural Networks (RNNs). RNN differs from Artificial Neural Network (ANN) by incorporating a feedback structure, allowing them to retain information from previous time steps and effectively process time-dependent data. RNNs are the basic method in this field, but gradient explosion and gradient disappearance problems make their training difficult (Ceni, 2025). In addition, due to these problems, RNNs have difficulty using system state information used for a long time. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been proposed to overcome these limitations. While LSTMs can preserve long-term information thanks to the memory cells and gate mechanisms in their structures, GRUs show similar performance with a simpler structure. The schematic representation of the structures of the methods is given in Fig. 2.

2.2.1. Long Short-Term Memory (LSTM)

LSTM is a special type of RNN designed to cope with the problems of RNNs even with long delays (Hochreiter and Schmidhuber, 1997). In this method, the long-term memory of RNNs is encoded in weights that change slowly during the training of the model, so the method is called LSTM. While RNNs in the LSTM structure maintain long-term memory, temporary activations exchanged between nodes provide short-term memory as shown in Fig. 2 (Waqas and Humphries, 2024). LSTMs consist of three gate mechanisms: Forget Gate, Input Gate, Output Gate. To enhance clarity, let the network input be $x_t \in R^{nxm}$ and the hidden units be h. Define n as the number of samples and m as the number of features. Let h_{t-1} represents the hidden state from the previous time step. The input gate is given by $I_t \in R^{nxh}$, the forget gate by $F_t \in R^{nxh}$, and the output gate by $O_t \in R^{nxh}$. The forget gate determines how much of the previous cell state will affect the current cell state. Input and output gates are the gates through which the cell is entered and exited. The equations for the gates are as follows (see Eq.s (5), (6) and (9)):

$$F_t = sigmoid(x_t W_f + h_{t-1} U_f + b_f) \tag{5}$$

The sigmoid layer of the input gate determines which information needs to update as follows.

$$I_t = sigmoid(x_t W_i + h_{t-1} U_i + b_i) \tag{6}$$

The *tanh* layer of the input layer creates a new candidate vector (C_t'). The current cell state is updated as follows:

$$C_t' = \tanh(x_t W_c + h_{t-1} U_c + b_c) \tag{7}$$

$$C_t = i_t C_t' + f_t\, C_{t-1} \tag{8}$$

In the last step, the output gate value of LSTM is calculated as follows:

$$O_t = sigmoid(x_t W_o + h_{t-1} U_o + b_o) \tag{9}$$

W_i, W_o, W_f , W_c and U_i, U_o, U_f , U_c are weights of the neural network. b_i, b_o, b_f , b_c are biases. LSTMs differ from RNNs with the gates they use in hidden states. These gates solve the problems in RNNs by deciding when to update the hidden state. The hidden state update is as follows:

$$h_t = O_t tanh(C_t) \tag{10}$$

2.2.2. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) introduced in (Cho et al., 2014) makes predictions by combining the information from previous time steps with the current time information. The GRU structure is shown in Fig. 2. The GRU architecture is similar to the LSTM architecture. These networks consist of gates and the information flow is controlled using these gates. The gate architecture allows the model to recognize temporal

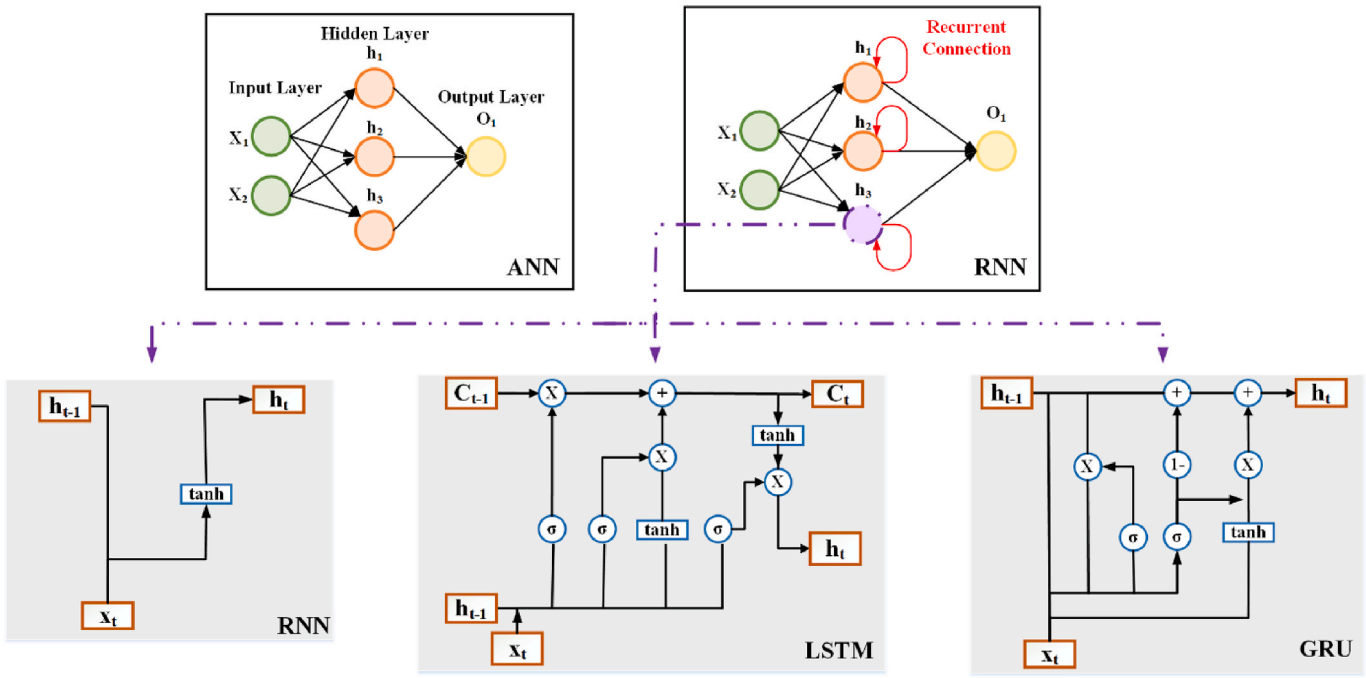

Fig. 2. Schematic representation of Artificial Neural Network (ANN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and GRU architectures.

relationships and manage long-term memory (Greff et al., 2016). In GRU, the gate architecture of LSTM is made simpler and more efficient. While LSTM uses three main gates, namely input, forget and output gates, GRU combines the input and forget gates to form the update gate and associates the output gate directly with the hidden state. Reset gate controls how much of the previous memory information is forgotten. The update gate (z_t) and reset gate (r_t) are obtained as follows:

$$z_t = sigmoid(W_z * (x_t, h_{t-1})) \tag{11}$$

$$r_t = sigmoid(W_r * (x_t, h_{t-1})) \tag{12}$$

where t is the time index. x_t is the input signal at time t. h_{t-1} is the prior hidden state. W_z and W_r are weight matrix of the update and reset gates, respectively. While calculating the candidate's hidden layer as shown in Eq. (13), past time information is preserved. The past time information is controlled by adjusting the $r(t)$ value.

$$h'_t = \tanh(W * (r_t * h_{t-1}, x_t)) \tag{13}$$

where W is the weight matrix. In the last step, the final hidden state is calculated by combining the previous hidden state ($*h_{t-1}$) with the new candidate hidden state (h'_t) as follows:

$$h_t = (1 - z_t) * h_{t-1} + z_t * h'_t \tag{14}$$

2.3. Multi-step feature forecasting approach

The simulation phase includes three distinct scenarios, designed to account for the structural differences in the model used during the input feature estimation phase. The series of independent variables is first decomposed into different frequency components using EMD, aiming to extract meaningful features from the time series. The initial decomposition identifies the most rapidly changing fluctuations and noise, with the first component typically containing random fluctuations, high-frequency noise, and short-term variations. To further refine the frequency components, the IMF-worst (high-frequency signal) undergoes VMD decomposition. This step enhances the model's ability to capture finer frequency details hidden within the IMF-worst. Finally, deep learning methods are applied to the resulting IMFs, and the predicted value is obtained by aggregating the outputs from each decomposition step. The parameters of the applied deep learning methods are optimized using the Bayesian tuning method. The output of this stage is used as input for the subsequent electricity price prediction model in the specific bidding zone. The details of three different scenarios, in which the methods used at this stage are analyzed based on their structural differences, are presented below.

2.3.1. Scenario 1: EMD + LSTM

In the first scenario, a single-step decomposition is performed, where EMD is applied to extract IMFs with different frequency components. Each IMF is then individually predicted using LSTM networks, and the results are combined to generate the final forecast as shown in Fig. 3. This approach enhances the capture of diverse frequency components in the time series and improves forecasting accuracy by leveraging deep learning models. In this scenario, separate forecasting models are employed for each component (sub-series) extracted through EMD. Each component is individually forecasted and later aggregated to generate the final prediction. After that, LSTM is applied to each component, which has been separated into different frequency ranges using the decomposition method, and the individual forecasts are aggregated to obtain the final prediction.

2.3.2. Scenario 2: EMD + VMD + LSTM

In this variant, a dual-stage decomposition is applied. In the first stage, N intrinsic mode functions (IMFs) are extracted based on the signal's frequency structure. After applying decomposition using EMD, the highest frequency IMF (IMF_1) primarily contains noise and short-term variations. In the second stage, the most complex and variable IMF (IMF_1) extracted by EMD is further decomposed using VMD, resulting in K new variational mode functions (VMFs) as shown in Fig. 4. By leveraging the smoothing filter capability of VMD, noise is effectively isolated and removed, enhancing the clarity and predictability of the decomposed components (Chen et al., 2017). All extracted IMFs and VMFs are individually predicted using the LSTM model to capture their temporal patterns and generate accurate forecasts.

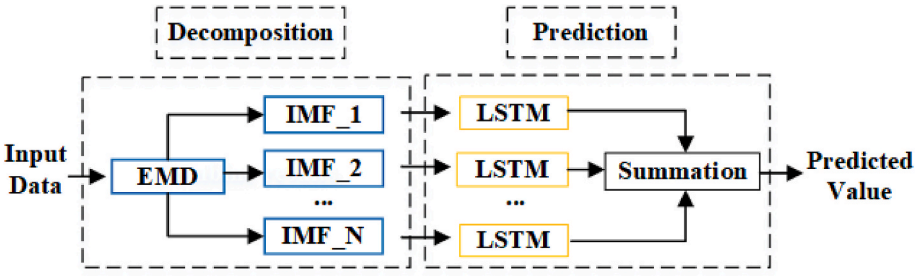

Fig. 3. EMD-based decomposition and LSTM forecasting framework.

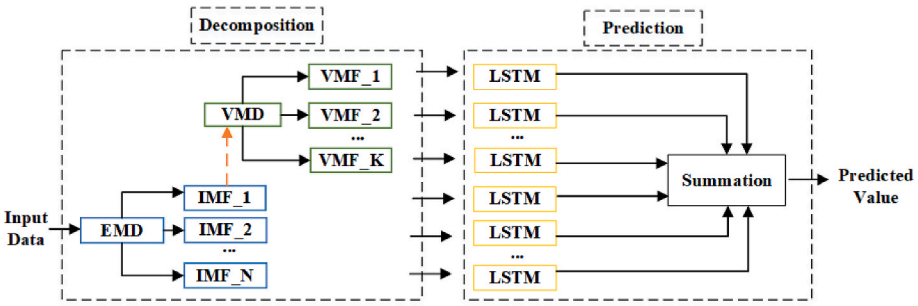

Fig. 4. Hybrid EMD-VMD decomposition and LSTM forecasting framework.

2.3.3. Scenario 3: EMD + VMD + LSTM + GRU

In this model, different prediction models are applied for high frequency components. The low-frequency components obtained from EMD decomposition are used for longer-term predictions and usually have less uncertainty. The highest frequency IMF (IMF_1), which contains noise and short-term variations, is further decomposed using VMD in this step. Subsequently, GRU is employed to predict each VMF signal as shown in Fig. 5. By analyzing dependencies and variable interactions over time, GRU demonstrates strong forecasting performance. Notably, GRU has been observed to outperform the LSTM method when dealing with noisy data (Qi et al., 2023).

2.4. Day-ahead electricity price prediction model

The estimated values from the previous stage are used as inputs to the price forecasting model in this stage. The model is trained on these inputs to learn the underlying relationships. As a result, it generates forecasts for all 24 h of the next day, producing a complete set of hourly electricity price estimates.

As mentioned earlier, in this stage, the model's aim is to make the day-ahead (H = 24) electricity price forecast for day D, i.e. $Y_D = \left(Y_{D,1}, Y_{D,2}, \cdots, Y_{D,H}\right)$. The following 62 input variables are used in the estimation models:

- Past day-ahead prices for a previous day, i.e. $Y_{D-1} = \left(Y_{(D-1),1}, Y_{(D-1),2}, \cdots, Y_{(D-1),H}\right)$,
- Total load forecasts for the current day, $X_D^L = \left(X_{D,1}^L, X_{D,2}^L, \cdots, X_{D,H}^L\right)$,
- Electricity generation (at D and $H = 1$) from
  - o Thermal energy sources, $X_{D,H}^{TE}$,
  - o Hydropower sources, $X_{D,H}^{HP}$,
  - o Solar sources, $X_{D,H}^{S}$,
  - o Onshore wind power sources, $X_{D,H}^{ONW}$,
  - o Offshore wind power sources, $X_{D,H}^{OFW}$,
  - o Other renewable sources, $X_{D,H}^{RS}$,
- Amount of the electricity exchange amounts (at D and $H = 1$), $X_{D,H}^{EX}$,
- Dummy variables representing the day of the week, X_D^{DoW} for D = 1,2, ⋯,7, here each variable specifies a specific day with a binary value.

Among the aforementioned inputs, past day-ahead prices (Y_{D-1}) and total load variables (X_D^L) have a dimension of (H = 24), while the remaining variables ($X_{D,H}^{TE}, X_{D,H}^{HP}, X_{D,H}^{S}, X_{D,H}^{ONW}, X_{D,H}^{OFW}, X_{D,H}^{RS}, X_{D,H}^{EX}$) are represented as single values. In addition, seven binary dummy variables representing the day of the week (X_D^{DoW}).

The first hourly power generation forecasts for the following day can be predicted one day in advance by system operators or models. To ensure the model remains realistic in terms of data availability, these

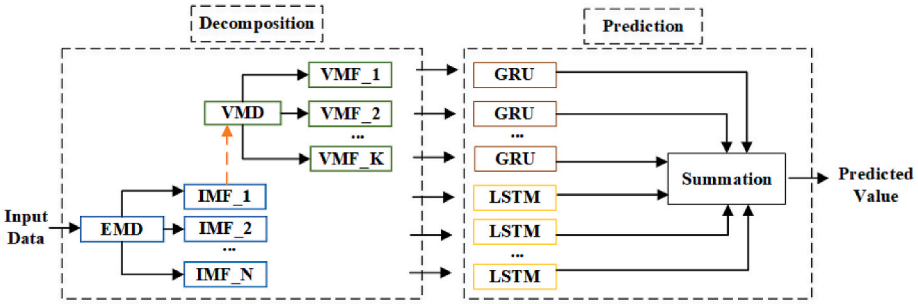

Fig. 5. Hybrid EMD-VMD decomposition with LSTM-GRU forecasting framework.

initial forecasts are preferred. In addition, using only a single time point instead of taking all 24-h energy production series as input reduces the input size of the model and provides faster learning. In addition, since spot prices are directly related to production capacity and supply-demand balance, especially in the first hours of the day, the input data of the model is organized in this way in this study. Then, a deep learning model with an LSTM-GRU hybrid architecture is developed to predict day-ahead electricity prices over a 24-h horizon. The model adopts a decentralized architecture in which each feature is processed individually through separate LSTM and GRU layers. The outputs of these layers, representing the learned patterns for each feature, are then concatenated to form a unified feature representation. This structure allows the model to capture feature-specific temporal dynamics more effectively. Finally, the combined vector is passed through a two-layer dense network to generate the 24-h price forecasts.

2.5. Introduction of hyperparameter optimization model in the developed model

In the data processing methods used, selecting the hyperparameters of the model is the principal disadvantage. Comprehensive search methods such as Grid Search (GS) and Random Search (RS) have been used determining the optimal hyperparameters. However, these methods have disadvantages such as computational cost, stochastic structures and long running times (Hanifi et al., 2024). In these methods, hyperparameter settings are considered independently of each other. Instead, using the information from previous parameter trials in the improvement of the next search space can provide improvements in terms of time and cost. For this reason, probabilistic methods are suggested for hyperparameter optimization. A goal of optimization is to find a point that minimizes a problem's objective function. In this field, Bayesian Optimization Algorithm (BOA) enables the search process to be completed efficiently using the defined information. At each step, it makes new estimates using the information it has previously obtained (Eleftheriadis et al., 2024).

The Bayesian Optimization method is based on the Bayesian Gauss Theorem and uses an acquisition function when evaluating the next hyperparameter value. The Bayesian optimization problem is expressed as follows:

$$x^* = arg\ min_{x \epsilon R} f(x) \tag{15}$$

Where x^* represents the optimized hyperparameter combination. R and $f(x)$ denote set of hyperparameter and objective function, respectively. In this work, the objective function is defined as the error of the results obtained from the prediction model with respect to the actual values.

In this study, the Bayesian Optimization method was used to search the parameter values of the estimation methods. Possible parameter values were selected from predefined value ranges. The Gaussian Process model evaluates each parameter combination, makes the best estimate based on the results of the tested parameters and determines which parameter combination should be tested in the next step. In this process, it is aimed to optimize the parameters that will give the best result among certain values. As a result, the performance of the model is improved by finding the most appropriate combinations on the discrete parameter space. The search values of the prediction models used in the study are given in Table 1.

Table 1
Optimized hyper-parameters and value ranges in deep learning algorithms.

Hyper-parameters	Value range
Number of LSTM units	(32, 128), step = 16
Number of GRU units	(32, 128), step = 16
Number of dense units	(64, 256), step = 32
Dropout rate	(0.1, 0.5), step = 0.05
Learning rate	Log-scale [0.0001–0.03]
Batch size	{16, 32, 64}

2.6. Performance metrics

To evaluate the forecasting performance of the methods, four error criteria are used as given in Table 2.

In addition, in this work, the significance testing has been done using the Diebold-Mariano (DM) test which is a popular paired comparison method (Diebold and Mariano, 1995) to compare the forecasting performance of two time series models and determine which model's forecasts are superior. One of the advantages of this test is that it can be used with different loss functions. This feature allows for a more comprehensive comparison between forecasting models.

The DM test was designed to compare two time series forecasts based on a user-defined loss metric. The difference in loss, d_t, between the two forecasts is calculated as:

$$d_t = L(error_{g,t}) - L(error_{h,t}) \tag{16}$$

Errors calculated at time t for model g and model h are represented as $error_{g,t}$ and $error_{h,t}$ and loss function is given as $L(.)$. The average loss is calculated as:

$$\overline{d} = \frac{1}{T}\sum_{t \epsilon T} d_t \tag{17}$$

And the DM test statistic is calculated as follows:

$$DM\ statistic = \frac{\overline{d}}{\sqrt{Var(\overline{d})}} \tag{18}$$

where, $Var(d)$ is the variance of the mean loss differential. A typical normal distribution is asymptotically followed by the DM statistic, meaning that when the sample size is large enough, the DM test statistic approaches the standard normal distribution. This property increases the reliability of the test and makes it easier to determine whether the performances of different prediction models are statistically significant. In this study, three different significance thresholds ($\alpha = 0.01, 0.05, 0.1$) were used to determine the significance level of the DM test. This indicates that hypothesis testing was conducted at different confidence levels, corresponding to 90 %, 95 %, and 99 % confidence intervals, respectively.

2.7. Case study

To meet its ambitious 2030 and 2050 carbon-neutral society goals, Denmark heavily relies on clean energy sources for electricity

Table 2
Error criteria and their equations.

Index	Abbreviator	Equation
Root Mean Square Error	RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{n \epsilon N}(x_{actual} - x_{predicted})^2}$
Mean Absolute Error	MAE	$MAE = \frac{1}{N}\sum_{n \epsilon N}\lvert x_{actual} - x_{predicted}\rvert$
Symmetric Mean Absolute Percentage Error	sMAPE	$sMAPE = \frac{1}{N}\sum_{n \epsilon N}\left\lvert\frac{x_{actual} - x_{predicted}}{(\lvert x_{actual}\rvert + \lvert x_{predicted}\rvert)/2}\right\rvert$
Relative Root Mean Square Error	rRMSE	$rRMSE = \sqrt{\frac{\frac{1}{N}\sum_{n \epsilon N}(x_{actual} - x_{predicted})^2}{\sum_{n \epsilon N}(x_{actual})^2}}$

Note that, x_{actual} and $x_{predicted}$ represent the real and predicted output values, respectively. N is the number of data points and n is the data index.

production, primarily wind turbines (WTs). In addition, the Danish electricity market is part of the European electricity market and operates within Nord Pool. The Danish power grid is divided into two main bidding zones: the Western grid (DK1 - Jutland and Funen region) and the Eastern grid (DK2 - Zealand region), which are interconnected via the Great Belt power connection.

In this study, an analysis of the DK1 bidding zone is conducted according to the available data for the period September 1, 2023–February 1, 2024. A proposed forecasting framework is applied to predict volatility and fluctuations in electricity prices by leveraging key market characteristics. The dataset used in this study includes historical electricity prices, total load, and electricity generation data from various renewable sources such as hydropower, solar, wind, and others for DK1. Furthermore, it incorporates electricity generation from thermal sources and electricity trade volumes with neighboring countries. The relevant data is available and can be sourced from (Energidataservice web page). The detailed information about the data set is given in Table 3.

2.7.1. Correlation analysis

As specified in Table 3, 14 features in the dataset are intended to be used as input features in electricity price forecasting. Therefore, Mutual Information and Spearman Correlation Analysis are utilized to examine the relationship between electricity price and these features. These methods can detect non-linear relationships, making them ideal for studying the connections between the target variable and input variables in detail. In this way, the accuracy of the selection of input variables used in our dataset is ensured. When the results obtained with different methods are analyzed, it is observed that the dependency values vary as seen in Figs. 6 and 7. This is because each method evaluates the relationships in the data set from different aspects and has different sensitivities.

2.7.2. Statistical analysis

The dataset has minimal missing values, making it consistent and well-maintained. Missing data were filled by averaging the previous and next feature values. Statistical Summary of energy market variables is presented in Table 4. Having such a wide range of electricity prices indicates the volatility of the market and underscore the importance of the accurate price prediction.

Table 3
Description of the input and target features.

Input variables	Description
Total Load	Total consumption including transmission loss (MWh)
Biomass	The average power generation from power plants where biomass is the primary fuel source (MWh)
Fossil gas	The average power generation from power plants where fossil gas is the primary fuel source (MWh)
Fossil hard coal	The average power generation from power plants where fossil hard coal is the primary fuel source (MWh)
Fossil oil	The average power generation from power plants where fossil oil is the primary fuel source (MWh)
Hydropower	The average power generation from hydroelectric power plants (MWh)
Other renewable	The average power generation from power plants where other fuels is the primary fuel source (MWh)
Waste	The average power generation from power plants where waste is the primary fuel source (MWh)
Solar power	Electricity generation from solar power plants (MWh)
Onshore wind power	Average electricity generation from onshore wind power (MWh)
Offshore wind power	Electricity generation from offshore wind power (MWh)
Exchange continent	It is the amount of exchange with Europe. Positive values indicate imports, negative values indicate exports (MWh)
Exchange Great Belt	This is the amount of trade through the Great Belt connection with DK2 region (MWh)
Exchange Nordic countries	It is the amount of exchange with Norway and Sweden (MWh)
SpotPrice in DK1	It refers to day-ahead spot prices for DK1 zone (MWh)

Having filled in the missing data, the next step is to decide on the input features to be used in the developed model. Considering the correlation values given in Figs. 6 and 7, the inputs to be used in the model are arranged. The variables for average electricity generation from power plants where biomass and fossil resources are the primary fuel sources are grouped together. Average electricity generation from power plants where other fuels and waste are the primary fuel source is combined and used as a single variable in the model. Amount of exchange with Norway and Sweden, Europe and DK2 region are also combined. While collecting the features, mean values were calculated to ensure consistency in the dataset. The Spearman results shown in Fig. 6, which capture ordered relationships, validate the accuracy of the combinations, particularly for renewable energy sources and electricity trade variables. Additionally, the Mutual Information results given in Fig. 7 confirm the strong linkage among thermal energy components as well as the other two combinations. In order to reduce the risk of overfitting and to ensure that the model focuses on fewer inputs, variables are grouped among themselves according to the results of correlation methods. Biomass, Fossil Gas, Fossil Hard Coal and Fossil Oil variables are grouped as Thermal Energy Sources, Other Renewable and Waste variables are grouped as Other Renewable Sources and Exchange Continent, Exchange Great Belt and Exchange Nordic Countries variables are grouped as Electricity Exchange. As a result, final 8 features are obtained to be used in the model.

A wide range of values can be found for the minimum and maximum values of different variables. Hence, when modeling, it is important to take into account these scale differences. The normalization process brings variables of different scales to the same level, which ensures that all inputs are equally focused. It also significantly reduces training time and improves convergence (Sola and Sevilla, 1997).

2.7.3. Data preprocessing

After data selection in the preprocessing phase, the next steps are the processing of null values and normalization. By completing these steps, the data to be used in the prediction phase is made suitable for the algorithm. In this study, the dataset contained a small amount of missing data. This missing data was filled by averaging the previous and next values.

Then, data normalization was performed to remove differences in scale in the dataset. This method is a widely used preprocessing step in machine learning and deep learning models. The specified data is scaled to the range [0, 1].

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{19}$$

In the final stage, to facilitate comparison with the original dataset, the prediction results are inversely normalized as follows:

$$x = x'(max(x) - min(x)) + min(x) \tag{20}$$

Where x and x' is input and output, respectively. min(x) denotes the lowest value for variable x, and max(x) refers the highest value for variable x.

3. Simulation and results

As mentioned earlier in the definition of the use case, the developed forecasting model is applied to the DK1 bidding zone. The dataset was split into approximately 80 % for training and 20 % for testing. A total of 2,928 data points from September 1, 2023, to December 31, 2023 is used as training set and of the remaining 744 data points from January 1, 2024, to January 31, 2024 is used as test set. The study consists of two stages. In the first stage, the input variables that influence the price of electricity are forecasted. After that, forecasting of prices for the day ahead is carried out. The pseudocode for the proposed framework is given in Appendix A. In addition, pseudocode for the multi-step feature

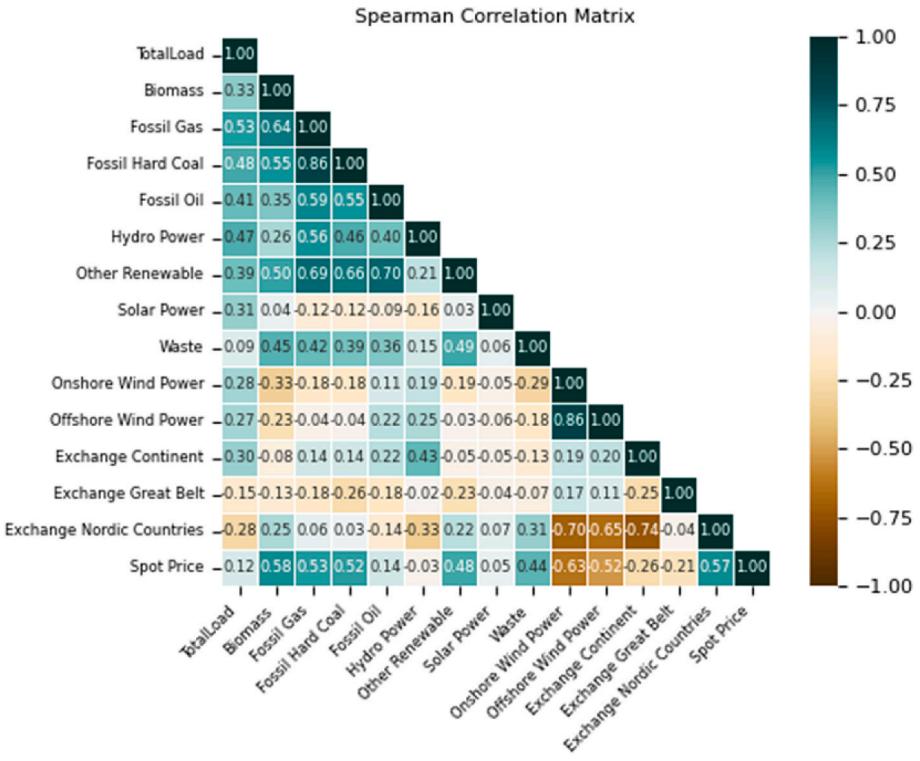

Fig. 6. Spearman correlation between the variables employed in the prediction models.

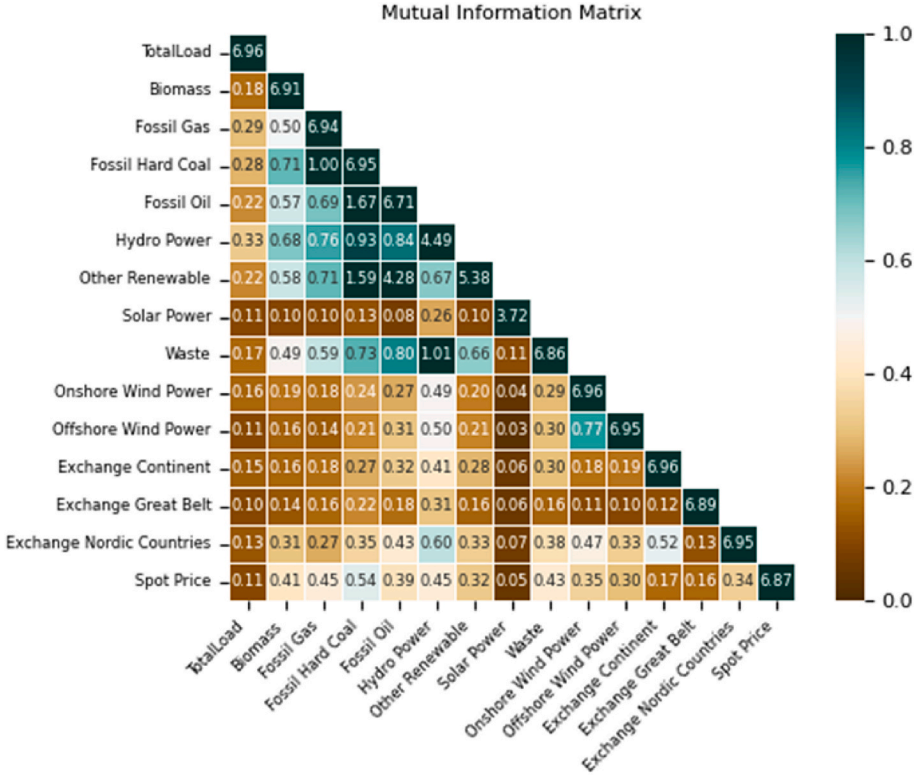

Fig. 7. Mutual Information correlation between the variables employed in the prediction models.

forecasting and pseudocode for the decentralized LSTM-GRU based day-ahead price forecasting framework is given in Appendix B and Appendix C, respectively.

This section consists of three subsections prepared to evaluate the effectiveness of the proposed model. In each subsection, the performance of the developed model is analyzed in comparison with existing approaches, and a comprehensive evaluation is presented on the basis of prediction accuracy and error metrics.

Table 4
Descriptive statistics of the input and target features between September 1st, 2023 and January 31st, 2024.

	count	mean	median	std	min	max
Total Load	3672	2646,65	2659,12	439,63	1477,89	3767,45
Biomass	3672	201,98	240,51	108,66	1,62	415,53
Fossil Gas	3672	174,11	129,34	118,36	37,55	482,02
Fossil Hard Coal	3672	321,57	211,57	282,4	34,86	1352,36
Fossil Oil	3672	22,49	15,37	12,72	5,38	70,61
Hydropower	3672	2,29	2,52	0,77	0,75	3,86
Other Renewable	3672	1,68	1,37	0,77	0,48	4,6
Solar Power	3672	126	0	290,17	0	1768,39
Waste	3672	44,35	46,2	16,17	8,09	111,73
Onshore Wind Power	3672	1258,94	1012,18	1007,83	3,44	3633,53
Offshore Wind Power	3672	689,43	718,89	410,08	0,83	1351,94
Exchange Continent	3671	−85,84	−360,06	1299,39	−3235,82	3299,8
Exchange Great Belt	3671	6,56	0	382,18	−590,85	600,32
Exchange Nordic Countries	3671	215,28	326,28	1491,44	−2377,36	2378,22
Spot Price (EUR) in DK1	3672	75,27	78,26	47,07	−8,54	524,27

3.1. Signal decomposition

Long-term trends and large-scale fluctuations in time series are captured within low-frequency components, while high-frequency components represent rapidly changing signals. Low-frequency components generally exhibit lower uncertainty, making them easier to forecast. In contrast, high-frequency components contain short-term fluctuations and noise, making direct estimation more challenging due to their chaotic nature as shown in Fig. 8. To improve predictability, an additional decomposition step is applied to further refine the high-frequency components as demonstrated in Fig. 9. In this step, three different scenarios are employed for forward-time forecasting of the features used in the prediction phase. Initially, time series are analyzed by decomposing them into low- and high-frequency components using EMD and the hybrid EMD-VMD approach. The original signals and the frequency components extracted through EMD decomposition are presented in Fig. 8.

Since the amount of electricity generated from solar power is directly influenced by the hourly variation in sunlight distinct daily periodic patterns are observed in this time series. In Scenarios 2 and 3, the re-decomposition of high-frequency components (IMF 1) disrupts these periodic patterns, adversely impacting the model's learning process. When IMF 1 is re-decomposed using VMD, the inherent periodic structure is altered, leading to increased instability in the forecast results. Consequently, further decomposition of components with short-term variability makes the forecasting model overly sensitive, resulting in higher error rates. Therefore, in the subsequent parts of the study, the results obtained from the Scenario 1 framework were utilized for the time series forecasting of electricity generation from solar power.

The EMD method applied in the first stage enables the time series to be decomposed into components based on their different frequency characteristics. Some of these components exhibit high noise levels and short-term variability, leading to reduced accuracy in the forecasting process. To address this issue, the high-frequency component is further decomposed using VMD. The graphs below (Fig. 9) provides a detailed illustration of the newly obtained components (excluding Solar Power Generation) after the VMD process.

The discrete and deterministic structure of the solar power series at different times of the day is shown in Fig. 10. Re-decomposition of high-frequency components can disrupt this regular structure in series with a distinctly periodic structure.

The decomposition structure in Scenario 3 disrupts the phase/energy consistency of the signal by breaking down predictable components. Therefore, the results from Scenario 1 were preferred for generation from solar power. Since this distinct periodic structure was not observed when examining the signals of the other features, Scenario 3 was successfully implemented.

At this stage of the study, different scenarios were proposed to demonstrate the success of the two-stage decomposition and the impact of using different deep learning methods based on signal structures. Therefore, it was decided to continue with Scenario 3 for the other features. In this study, the results obtained for each variable were examined to determine which scenario to preferred. Future work plans to develop an adaptive scenario selection mechanism that can automatically determine the most appropriate combination of decomposition and modeling based on the characteristics of each dataset.

3.2. Forecasting results and analysis

In this study, a decentralized architecture is implemented across three different scenarios. The intrinsic mode functions obtained from the decomposition phase are fed into separate forecasting models. Predictions are then generated under three distinct scenarios, and the results are evaluated based on various error metrics, as presented in Fig. 11. This evaluation includes four metrics: RMSE, sMAPE, MAE, and RRMSE. Additionally, to demonstrate the effectiveness of the decentralized architecture, the results from the centralized structure are also included for comparison.

As shown in Fig. 11, the decentralized architecture yields significantly better results. Comparing RMSE values reveals that scenario 2 achieves approximately a 50 % improvement in forecasting generation from other renewable sources compared to scenario 1. Moreover, predictions for load consumption, thermal energy generation, hydropower, and onshore wind improve by about 20 % on average. Moving from scenario 2 to scenario 3, further gains are observed: prediction accuracy improves by around 10 % for onshore wind, offshore wind, thermal energy, and other renewable sources, with smaller improvements in the remaining variables. Overall, scenario 2 outperforms Scenario 1, while Scenario 3 delivers the highest predictive accuracy across all criteria.

Among the three different scenarios, Scenario 3 demonstrates superior performance, achieving the lowest error rates across multiple error metrics. Its success can be attributed to the hybrid decomposition and forecasting approach, which enables a more effective analysis of various frequency components within the time series. By accurately capturing long-term trends in low-frequency components and enhancing the prediction of short-term fluctuations in high-frequency components, this framework significantly improves overall forecasting accuracy. As a result, a comprehensive analysis of the forecast results obtained using scenario 3 for seven different variables and scenario 1 for solar power generation is presented in Fig. 12.

As seen, the model's predictions are compared with actual values for each variable, and performance is evaluated using various error metrics to assess the accuracy and reliability of the forecasts. Furthermore, the performance differences between the centralized and decentralized architectures are highlighted by comparing the forecast results with those of the centralized model. By processing the various components of the

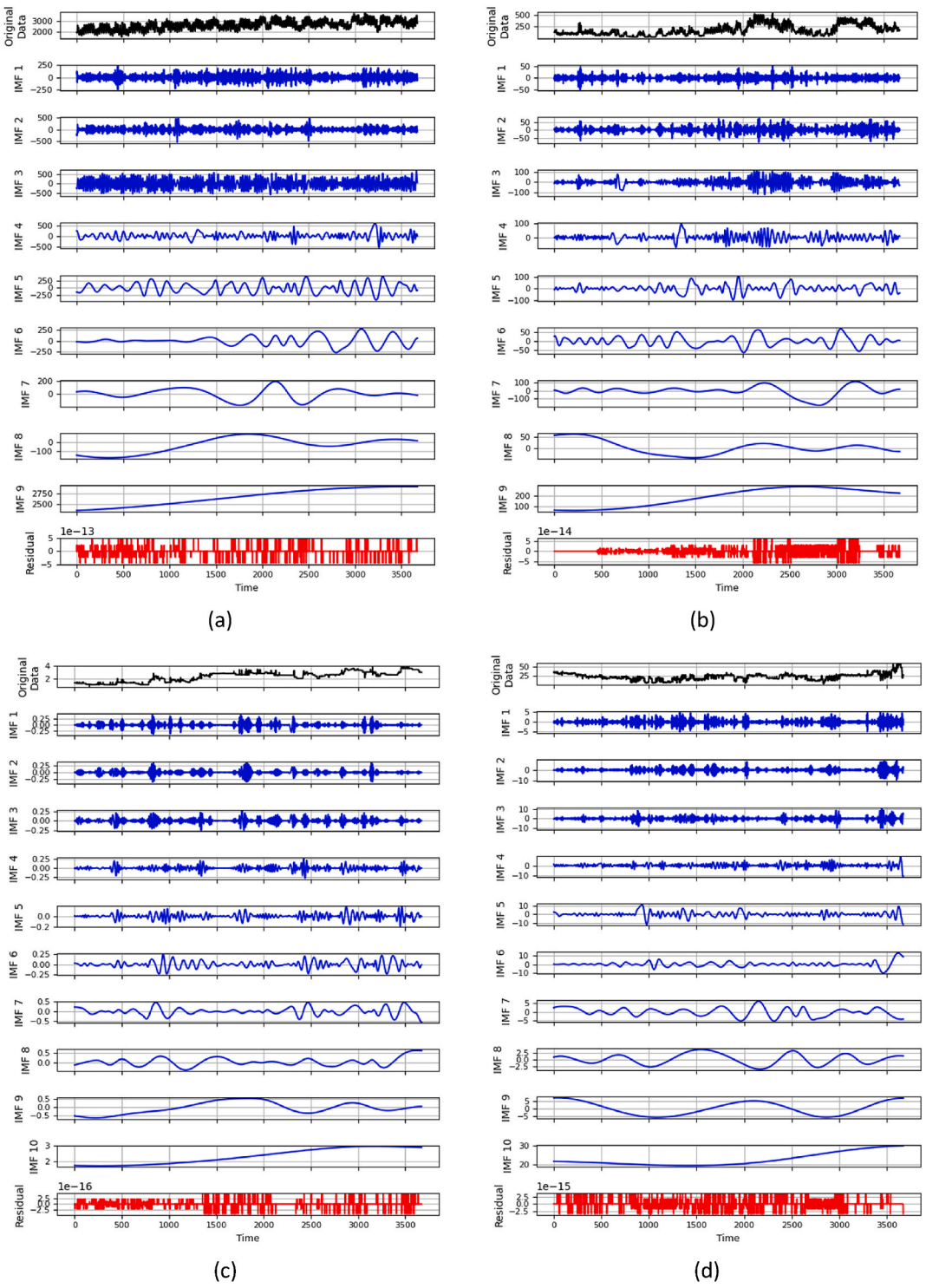

Fig. 8. Extraction of features based with EMD method (a)Load Consumption, (b)Generation from Thermal Energy Sources, (c)Generation from Hydropower, (d) Generation from Other Renewable Sources, (e)Generation from Solar Power, (f)Generation from Onshore Wind Power, (g)Generation from Offshore Wind Power, (h) Amount of Electricity Exchange.

time series independently, the decentralized architecture achieves lower error rates and greater reliability.

3.3. Prediction of day-ahead electricity price in DK1

Electricity price time series forecasting plays a crucial role in the energy market. However, complex characteristics such as high frequency, nonlinearity, and volatility make this forecasting challenging. Therefore, this complex structure needs to be better modeled with advanced time series forecasting methods.

In this study, electricity prices were forecasted one day in advance using multiple prediction scenarios and relevant input factors as discussed previously. Machine learning and deep learning-based models provide powerful tools for capturing the non-linear nature of electricity

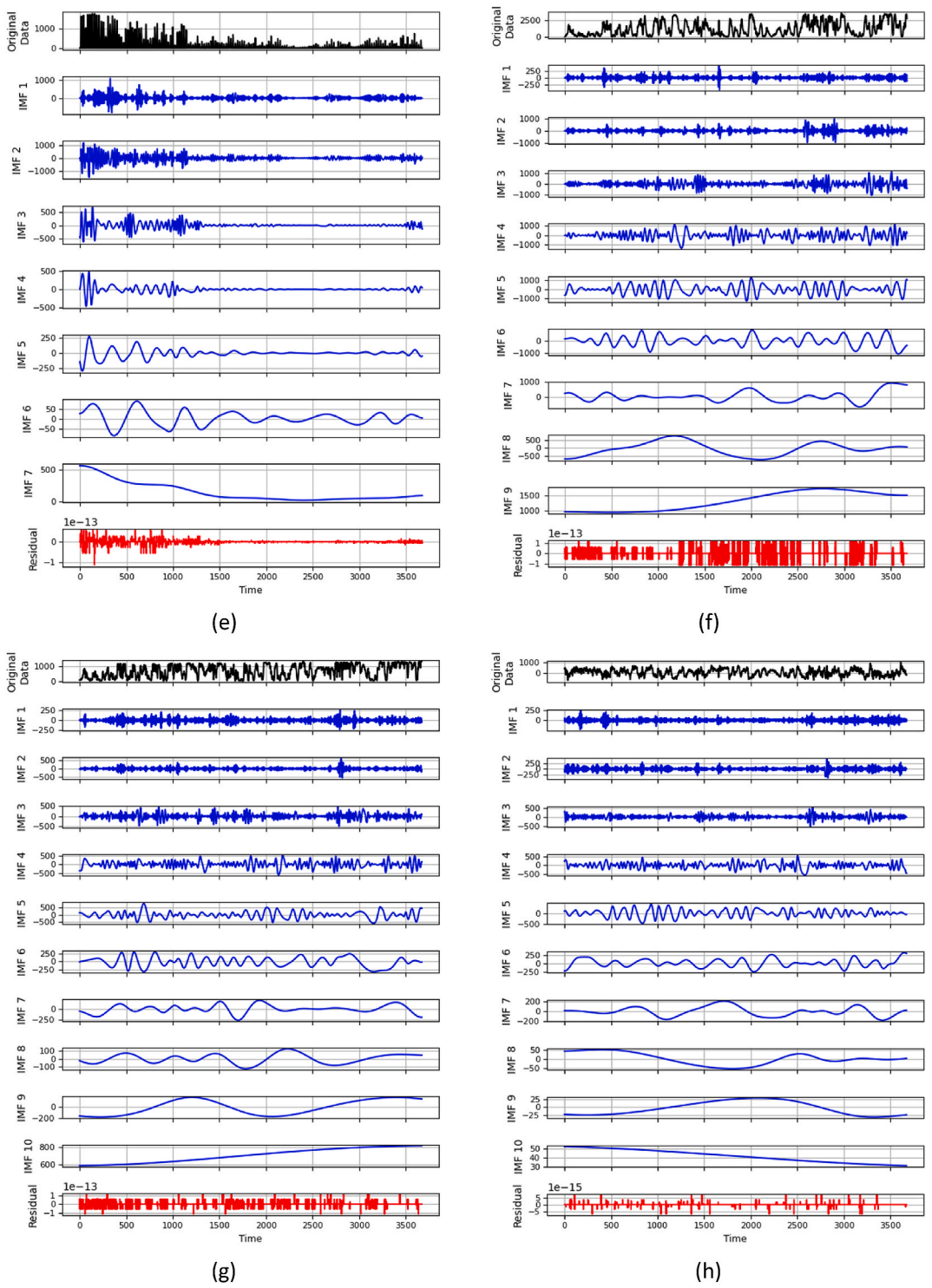

Fig. 8. (*continued*).

prices and accurately forecasting price fluctuations. At this stage of the study, day-ahead electricity price forecasting leverages feature values obtained from various scenarios to achieve the most precise price predictions possible. Fig. 13 compares the day-ahead electricity price forecasts obtained using different forecasting methods with feature values derived from multi-step forecasting. The black line represents the actual electricity prices, while the predictions from different models are shown with various colored lines. In general, all models follow the overall trend of real electricity prices, but their forecasting accuracy varies. A closer examination of the January 15–16 period reveals that some models align more closely with actual values than others, particularly in capturing peaks and troughs, which highlights their ability to represent fine-grained price fluctuations.

Table 5 presents the error metrics for different forecasting models, providing a numerical comparison of their performance. The results clearly indicate that decentralized models outperform centralized ones. Among them, the decentralizedLSTMGRU model achieves the lowest RMSE (22.55), SMAPE (28.70), MAE (18.11), and RRMSE (30.19), confirming that it delivers the most accurate forecasts. This suggests that decentralization plays a key role in improving predictive performance.

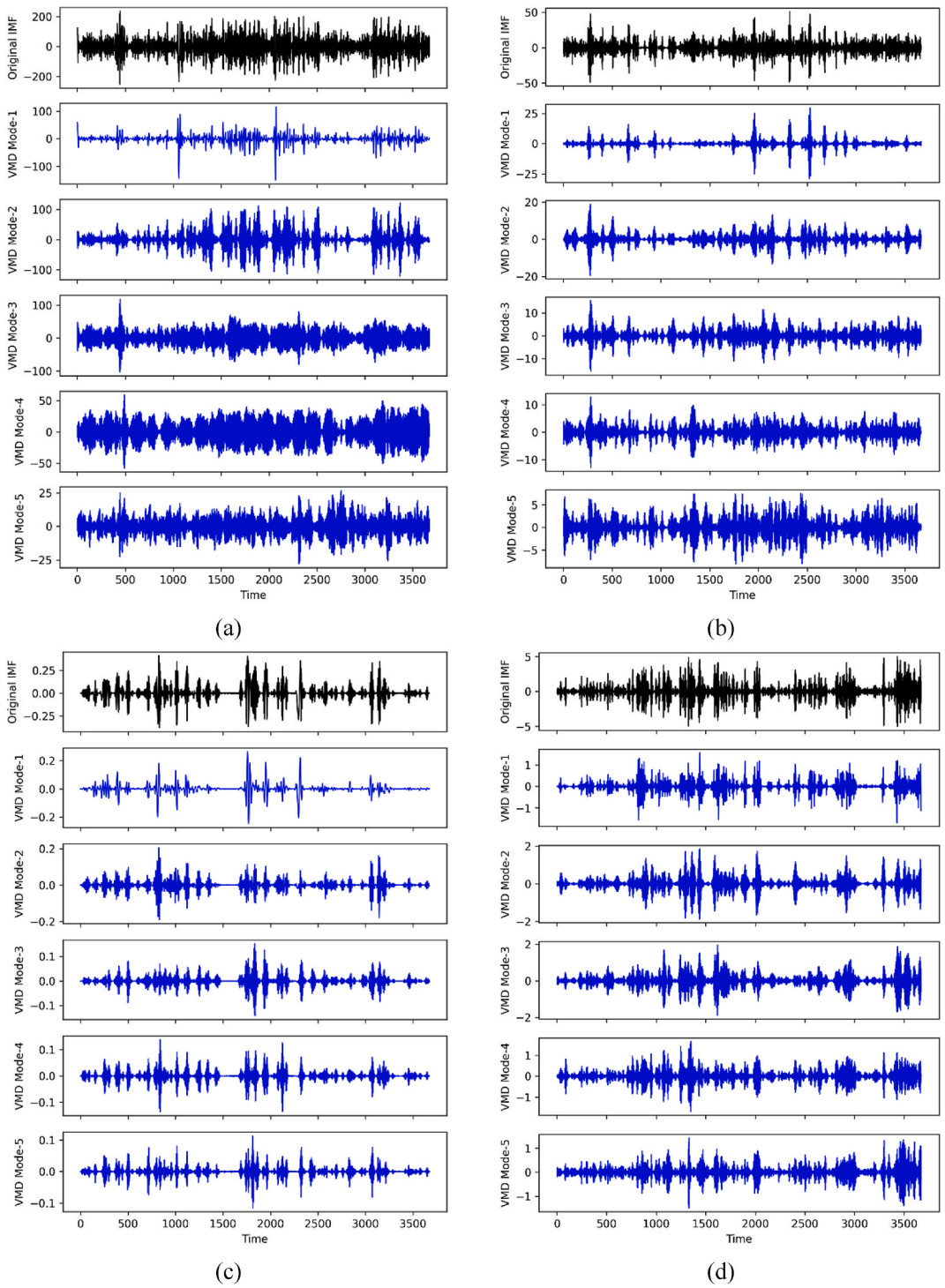

Fig. 9. Decomposition results of IMF-worst signal with VMD method (a)Load consumption, (b)Generation from thermal energy sources, (c)Generation from Hydropower, (d)Generation from other renewable sources, (e)Generation from onshore wind power, (f)Generation from offshore wind power, (g)Amount of electricity exchange.

When comparing centralized and decentralized deep learning models, it is evident that decentralized approaches consistently yield superior results. While traditional machine learning models perform reasonably well, they are outperformed by deep learning models. Despite their robustness, centralized deep learning models struggle to fully capture the complexities of price fluctuations, leading to higher error rates. In conclusion, decentralized models, particularly LSTMGRU, effectively minimize forecasting errors and enhance predictive efficiency, making them highly suitable for real-world electricity price forecasting. Their improved accuracy and ability to handle intricate market dynamics suggest strong potential for future applications in energy market analysis and decision-making.

Residual graphs which illustrate the residual (error) values between the actual and predicted values of various prediction models are shown in Fig. 14.

Each method is illustrated with a separate residual graph for clear

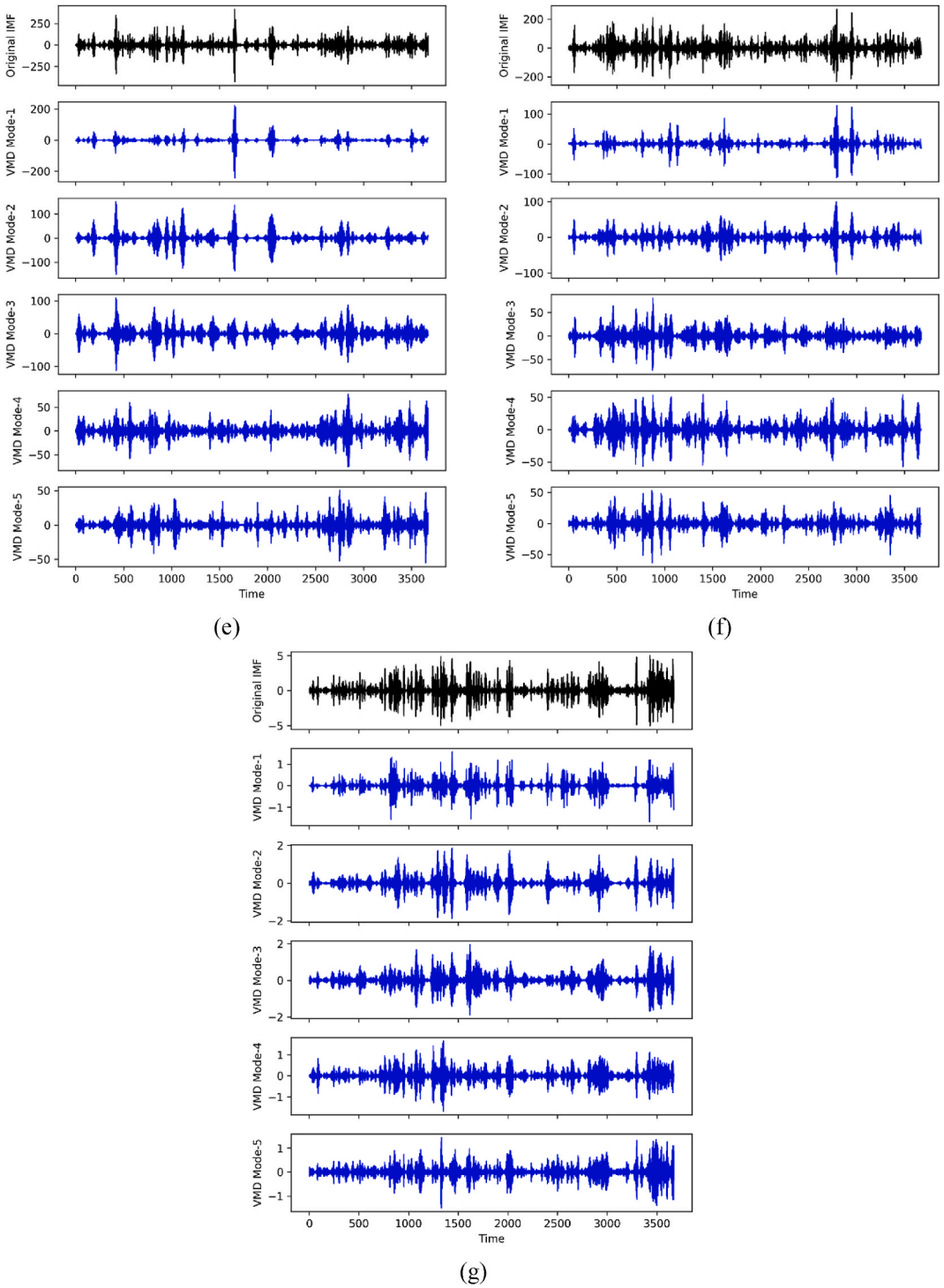

Fig. 9. (*continued*).

comparison. Although most models show residuals centered around zero, some exhibit more extreme errors. The residuals of the machine learning model (Fig. 14a–c) are more dispersed, with clusters of errors in specific regions. In the LGBMRegressor method results, residuals are moderately spread, but positive deviations increase as higher true values increase. CatBoost exhibits a similar pattern to LGBMRegressor, but the negative deviations are slightly less. In SVR, the majority of residuals are between −75 and + 75. However, while it appears more balanced at lower values, the model exhibits more positive deviations at higher values. Among them, the SVR model displays the widest error spread, indicating higher variance.

When we examine centralized structures, we see that errors are distributed over a wider range in CentralizedLSTM (−150 to +100). In CentralizedLSTM2, residuals are slightly more densely distributed and closer to the center than in the previous model. CentralizedLSTMGRU also has a generally quite scattered structure.

In contrast, the centralized (Fig. 14d–f) and decentralized deep learning models (Fig. 14g–i) show narrower error distributions, particularly in the decentralized case, which demonstrates fewer and less severe residuals. This suggests that decentralization improves learning

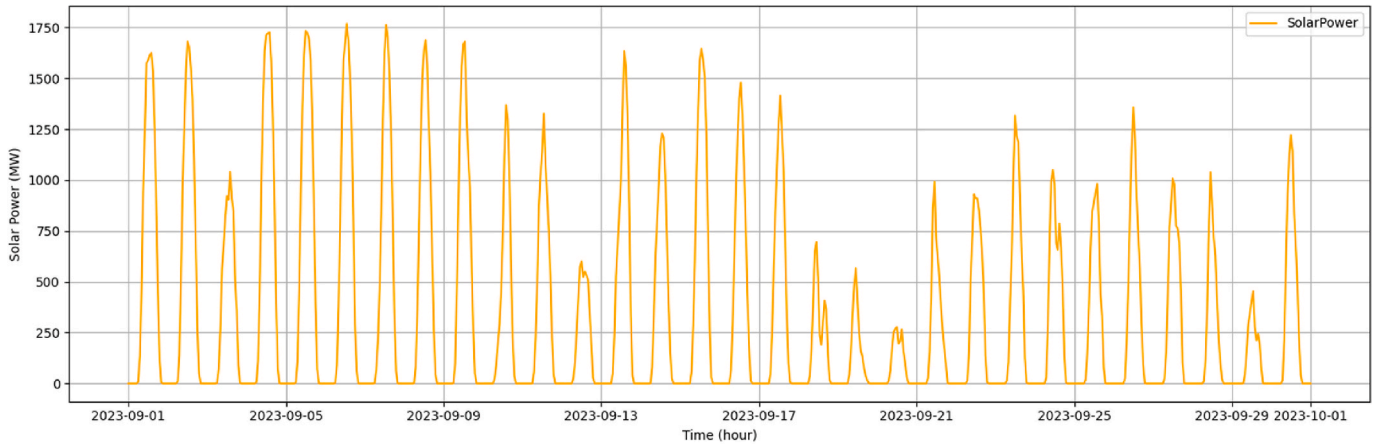

Fig. 10. Time series of solar power generation over time.

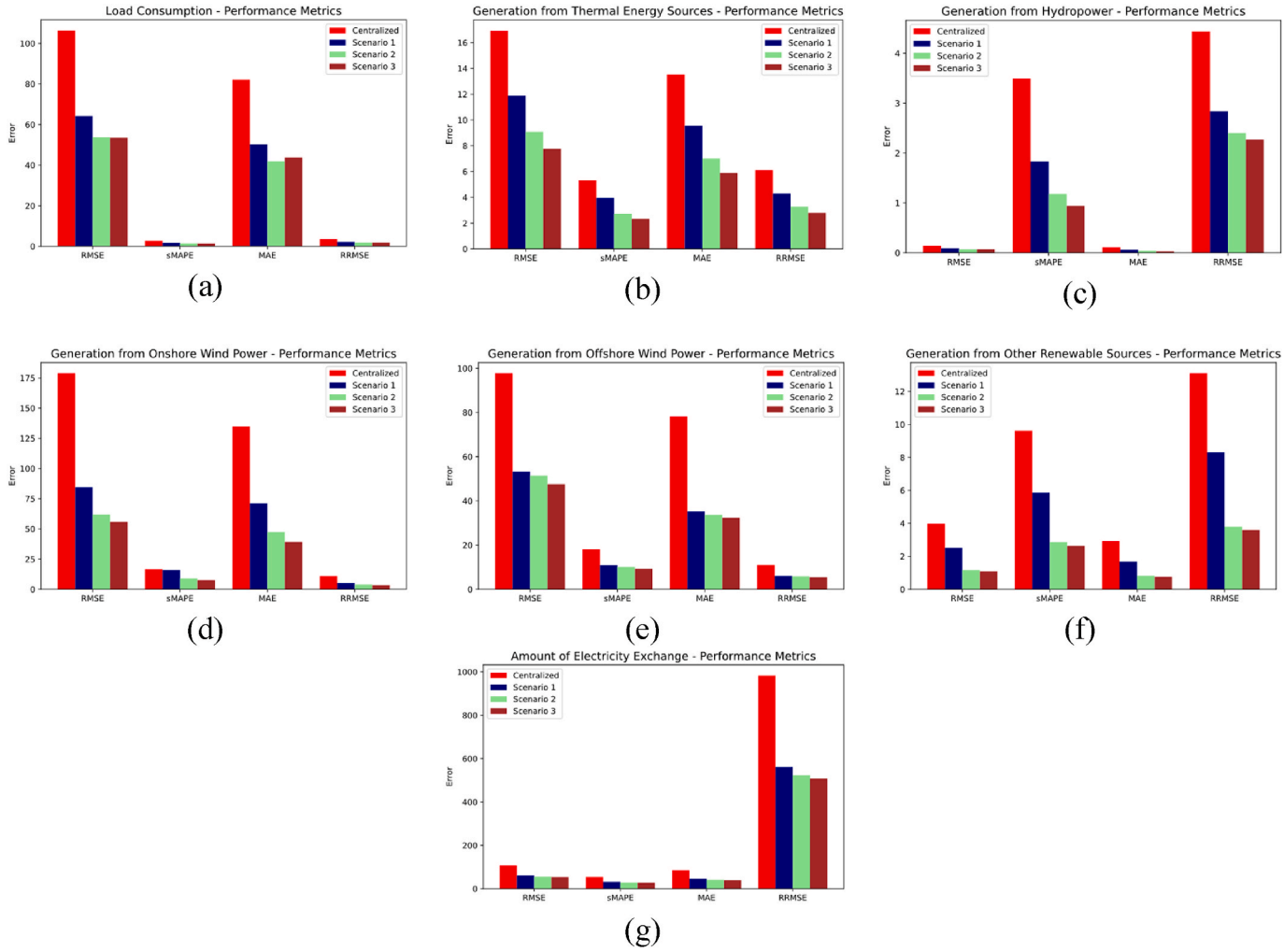

Fig. 11. Comparison of different decomposition scenarios across multiple error metrics (a)Load consumption, (b)Generation from thermal energy sources, (c) Generation from hydropower, (d)Generation from onshore wind power, (e)Generation from offshore wind power, (f)Generation from other renewable sources, (g) Amount of electricity exchange.

efficiency and model robustness. In DecentralizedLSTM, the errors are quite tightly distributed, between ±80 °C. A tighter distribution is visually observed. The DecentralizedLSTM2 results appear consistent with the model's overall predictive performance. The residual graph for DecentralizedLSTMGRU shows that this model has one of the most evenly distributed graphs among all other models. Errors generally fall between −60 and + 60, with the residuals being more concentrated around the center.

These results demonstrate that the model offers the most stable and predictable performance in terms of overall accuracy. Overall, the

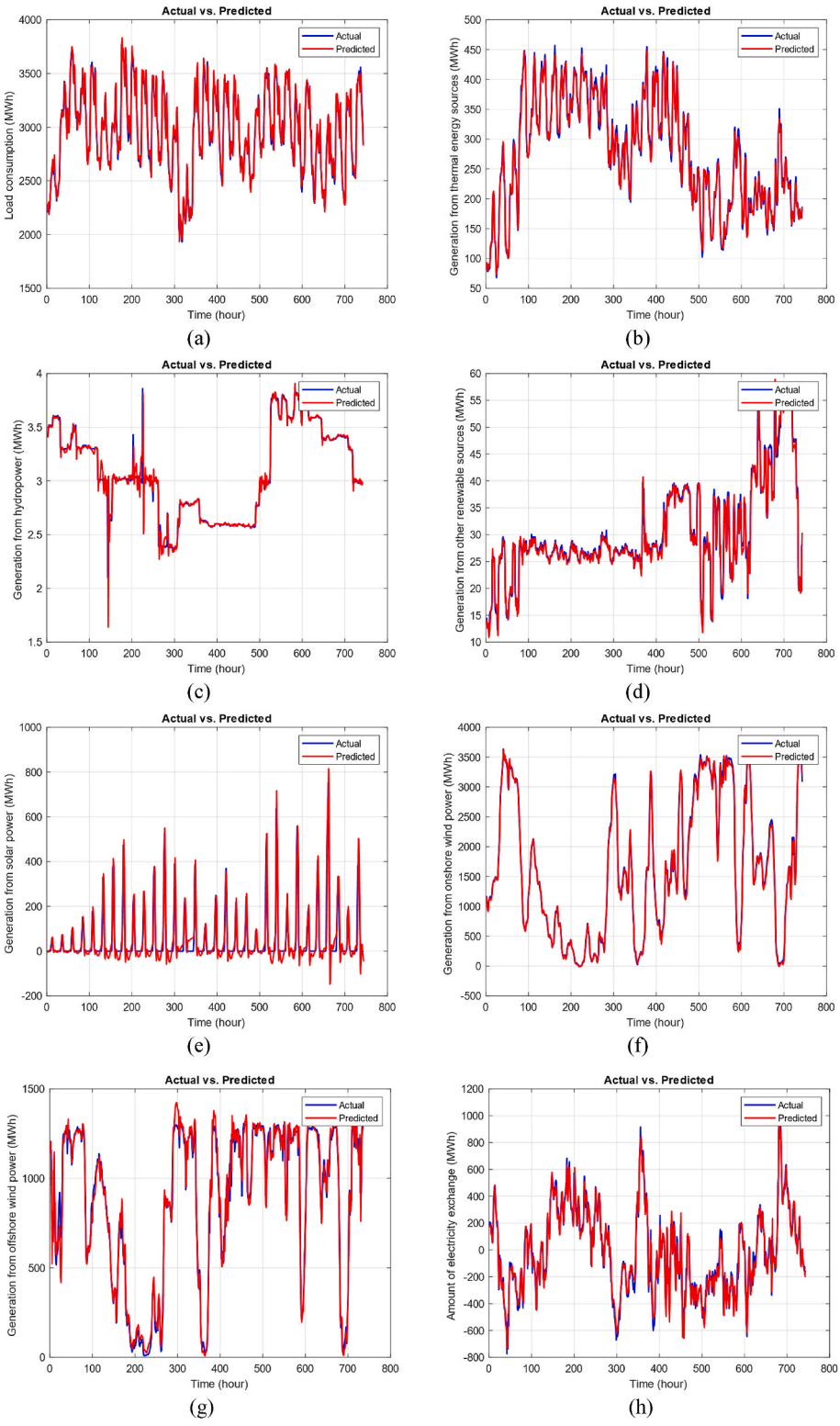

Fig. 12. Forecast and regression analysis of variables in Scenario 3. (a)Load consumption, (b) Generation from thermal energy sources, (c) Generation from hydropower, (d)Generation from other renewable sources, (e)Generation from solar power, (f)Generation from onshore wind power, (g)Generation from offshore wind power, (h)Amount of electricity exchange.

proposed model delivers the highest prediction accuracy, with the lowest error levels, outstripping all other approaches. Finally, the DM test was conducted to evaluate the prediction performance of the proposed model compared to other approaches. Three different loss functions (MSE and MAE) were used for evaluation as shown in Table 6.

In this test, the p-value represents the statistical significance of the results, where a lower p-value indicates stronger statistical significance. To enhance the clarity of the table, asterisks are used to denote different significance levels. The key observation from the table is that the DM statistics are significantly positive, indicating that the proposed model outperforms the other approaches. The results show that the difference in forecast accuracy between the compared models is statistically

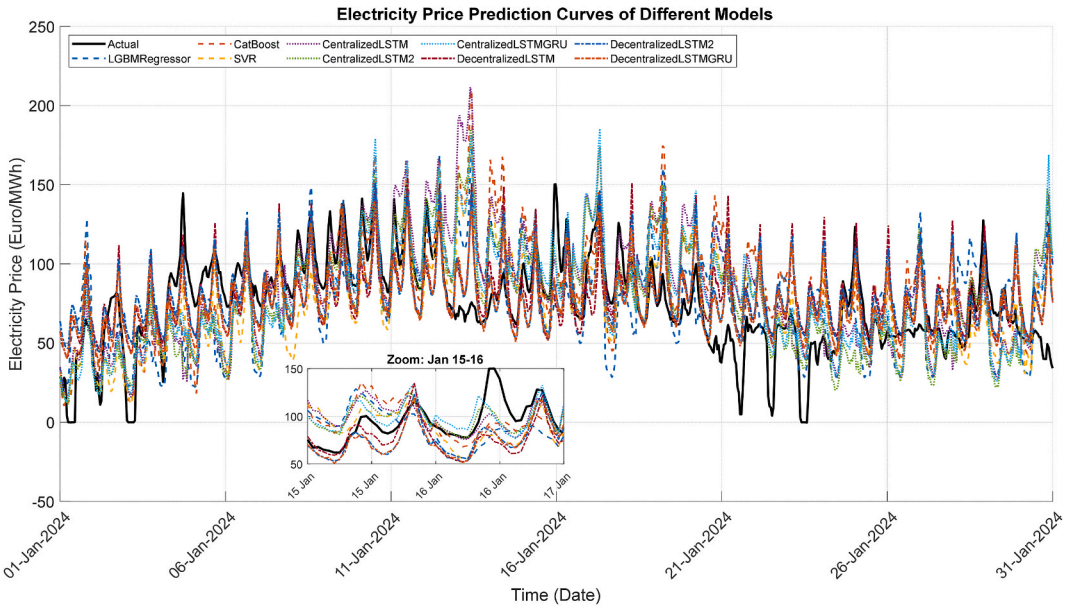

Fig. 13. Day-ahead price prediction result curve.

Table 5
Error metrics for different prediction models.

		RMSE	SMAPE	MAE	RRMSE
Machine Learning Methods	LGBMRegressor	31.38	40.15	25.53	42.01
	Catboost	31.97	36.29	24.65	42.81
	SVR	29.54	37.41	24.20	39.55
Deep Learning Methods	centralizedLSTM	33.93	37.85	25.86	45.43
	centralizedLSTM2	30.66	37.69	24.01	41.05
	centralizedLSTMGRU	29.68	36.34	23.75	39.74
Improved Deep Learning Methods	decentralizedLSTM	25.22	30.64	20.06	33.77
	decentralizedLSTM2	24.04	29.62	19.17	32.19
	decentralizedLSTMGRU	22.55	28.70	18.11	30.19

significant at the 1 % level, meaning that the null hypothesis of equal forecast performance is rejected with 99 % confidence. The results illustrated that this model achieves significantly better accuracy in daily price prediction for energy markets. Table 6 further highlights the advantages of decentralized resources by comparing them with centralized methods. As depicted, the decentralized structure effectively enhances prediction accuracy.

To conduct a sensitivity analysis to show how different input features impact forecasting accuracy, we performed feature ablation experiments by systematically removing each input and observing the corresponding changes in prediction performance. The basic logic in this method is to eliminate each feature or feature group and try different combinations of input features (Yang et al., 2025). Then, prediction performance results are examined to understand the impact of input features on prediction. The results of the sensitivity analysis showing the impact of removing each input feature on forecasting accuracy are presented in Table 7.

The results indicate that the removal of historical price information leads to the most significant performance degradation, highlighting it as the most critical variable for day-ahead electricity price forecasting. A moderate increase in forecasting error is observed when load consumption and day of the week features are excluded, suggesting their substantial but secondary influence. In contrast, the exclusion of features related to renewable energy generation results in comparatively smaller changes in model performance, indicating a relatively lower impact on price prediction.

4. Conclusion

In this study, a two-stage hybrid model was proposed for day-ahead electricity price forecasting with decomposition techniques. In the first stage, forward prediction of the features of electricity price data was made. In the second stage, the day-ahead price was forecasted using both the estimated features and historical price data. To handle the complex, multi-frequency structure of electricity price signals, the first stage employed EMD to decompose the signal into sub-series. The most complex high-frequency component is further decomposed using VMD. GRU networks were utilized for VMD decomposed signals (VMFs), while LSTM networks were employed for the prediction phase of EMD decomposed signals (IMFs). The t+1 predicted feature values and historical prices obtained from this step were then used in the second stage. In the second stage, a decentralized structure was adopted, and day-ahead forecasting was performed using a hybrid LSTM-GRU architecture. The model was designed to effectively capture the complex and high-frequency patterns commonly observed in electricity price data, and aimed to improve prediction accuracy through hybrid deep learning techniques. Bayesian optimization was applied for hyperparameter tuning to enhance model performance. Experimental results showed that the proposed method achieved significant improvements in forecasting accuracy. Compared to traditional machine learning models such as LGBMRegressor, CatBoost, and SVR, the proposed approach reduced RMSE by approximately 27.15 %, and compared to LSTM-based models, by approximately 28.24 %, demonstrating superior performance over benchmark methods.

In this study, some features were grouped for analysis. In future research, applying feature selection techniques may further enhance the model's overall accuracy and robustness. In the proposed hybrid discriminant-based prediction framework, sequential application of EMD and VMD signal decomposition techniques in the first step allows the prediction models to learn more detailed patterns, but it increases computational time and has a high computational cost. Therefore, the need for more processing power and time during the implementation phase may limit the usability of the models. Therefore, future work can explore strategies to reduce computational time. Furthermore, the sensitivity of deep learning models to hyperparameter tuning also has a significant impact on performance. While Bayesian optimization was used for hyperparameter tuning in this study, future work may explore the effectiveness of recently proposed parameter optimization

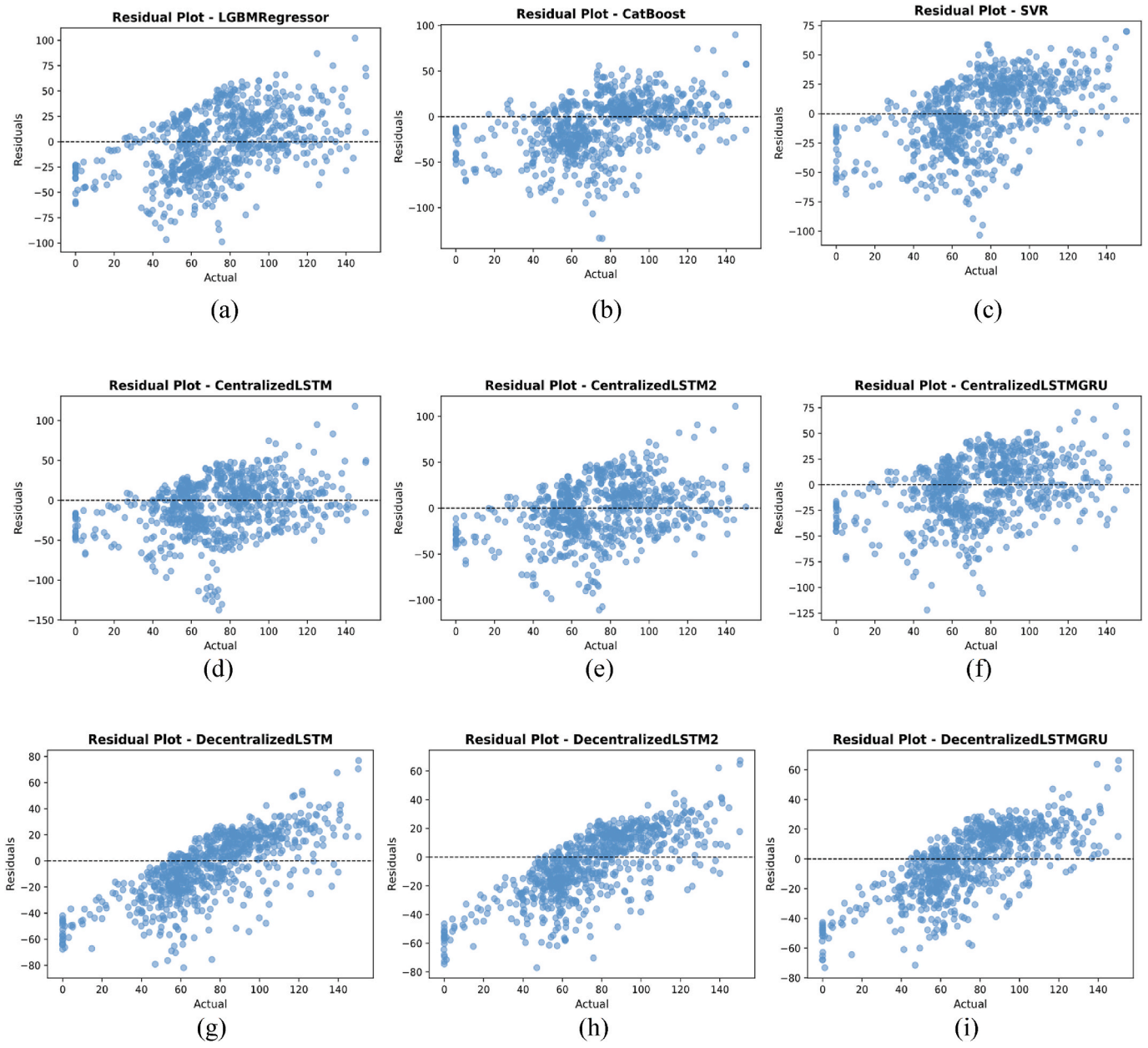

Fig. 14. Uniform residual distribution of forecasting models (a) LGBMRegressor, (b) CatBoost (c) SVR, (d) CentralizedLSTM, (e) CentralizedLSTM2 (f) CentralizedLSTMGRU, (g) DecentralizedLSTM, (h) DecentralizedLSTM2, (i) DecentralizedLSTMGRU.

Table 6
p-values of the Diebold and Mariano test with different criteria.

	MSE			MAE		
Compared Model	Loss_diff	DM value	p-value	Loss_diff	DM value	p-value
LGBMRegressor	476,229	10,719	0	7,42	10,508	0
CatBoost	513,68	8,598	0	6,536	9,116	0
SVR	364,02	9,846	0	6,087	10,921	0
CentralizedLSTM	642,655	8,336	0	7,75	9,413	0
CentralizedLSTM2	431,523	8,028	4E-15	5,895	8,16	2E-15
CentralizedLSTMGRU	372,396	8,322	0	5,641	8,535	0
DecentralizedLSTM	127,712	7,711	4E-14	1,948	7,292	7,82E-13
DecentralizedLSTM2	69,352	7,121	2,538E-12	1,064	5,838	7,9E-09

algorithms to further improve model performance.

The hybrid deep learning framework proposed in this study achieved high accuracy in predicting day-ahead electricity prices. Future work could integrate the proposed framework into market systems with real-time data streams, making it a secure tool for market participants and stakeholders to use in decision-making processes. Additionally, this study used data from the DK1 region of the Danish energy market. Future studies would be valuable in assessing the generalizability of the

Table 7
Ablation study of input features based on forecasting error metrics.

Feature Removed	RMSE	MAE	ΔRMSE
None (Full model)	22.55	18.11	–
Load Consumption	24.47	19.77	+1.92
Day of the Week	24.52	19.59	+1.97
Generation From Thermal Energy Sources	24.02	19.33	+1.47
Generation From Hydropower	24.18	19.46	+1.63
Generation From Other Renewable Sources	24.21	19.40	+1.66
Generation From Solar Power	23.54	18.95	+0.99
Generation From Onshore Wind Power	23.86	19.36	+1.31
Generation From Offshore Wind Power	23.43	18.95	+0.88
Amount Of Electricity Exchange	23.99	19.35	+1.44
Past Day-Ahead Prices	29.12	23.15	+6.57

model by applying it to energy markets in different geographic regions.

CRediT authorship contribution statement

İnayet Özge Aksu: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Sina Ghaemi:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Amjad Anvari-Moghaddam:** Writing – review & editing, Supervision, Investigation, Formal analysis.

Declaration of competing interest

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

Acknowledgments

İ.Ö.A. acknowledges support from the Scientific and Technological Research Council of Türkiye (TÜBİTAK) BIDEB-2219 Postdoctoral Research Fellowship Program and the BAP project No. 24142001, which was adopted by the Unit of Scientific Research Projects Coordination connected to Adana Alparslan Türkeş Science and Technology University.

Appendix A

The pseudocode for the proposed framework is given in Algorithm 1.

Algorithm 1. The pseudocode for the proposed framework

Input: Raw dataset: hourly time series data (Load Consumption, Generation Sources etc.)
- Hyperparameters for deep learning models
Output: Forecasted 24-h ahead LMP values: $\hat{y}\_\{T+1\} = [\hat{p}\_1, \hat{p}\_2, \ldots, \hat{p}\_\{24\}]$

Step 1: Preprocessing
- Handle missing values
- Normalize input features using MinMaxScaler
- Create lagged features for past day-ahead prices

Step 2: Feature Decomposition and Forecasting (Multi-Step Feature Forecasting)
For each selected input feature:
- Apply EMD to decompose the signal into IMFs: f → [IMF_1, …, IMF_n]
- Select the most complex IMF (e.g., IMF_ worst)
- Apply VMD to further decompose this IMF into K modes: IMF_ worst → [v_1, …, v_k]
- Group the resulting components into High-Frequency (HF) and Low-Frequency (LF) categories
- Train LSTM on LF components and GRU on HF components
- Forecast each component using the trained models
- Combine predictions to form multi-step forecasts for each feature: $\hat{f}\_\{T+1\}$

Step 3: LMP Prediction
- Collect all predicted input features $\hat{f}\_\{T+1\}$ from Step 2
- Concatenate as input for the final model
- Train hybrid deep model to predict day-ahead LMP
- Forecast day-ahead LMP using the trained deep model: $\hat{y}\_\{T+1\} \leftarrow$ Model($\hat{X}\_\{T+1\}$)
- Inverse transform predictions to original scale

Appendix B

The pseudocode for the Multi-Step Feature Forecasting is given in Algorithm 2.

Algorithm 2. The pseudocode for the Multi-Step Feature Forecasting

Input: Raw dataset: hourly time series data (Load Consumption, Generation Sources etc.)
- Hyperparameters for deep learning models
Output: Forecasted signal: 24-h ahead values $\hat{y}\_\{T+1\} = [\hat{y}\_1, \hat{y}\_2, \ldots, \hat{y}\_\{24\}]$

Step 1: Data Preprocessing
- Handle missing values and ensure numerical format
- Normalize the data using MinMaxScaler

(*continued on next page*)

(*continued*)

4	**Step 2**: Apply EMD Decomposition
5	- Decompose the time series into multiple IMFs: IMF_1, IMF_2, …, IMF_n
6	- Identify IMF_worst (highest frequency component)
7	**Step 3**: Apply VMD on IMF_worst
8	- Decompose IMF_worst into into K modes: IMF_ worst → [v_1, …, v_k]
9	**Step 4**: Deep Learning-Based Component Forecasting
10	For each IMF from EMD:
11	- Scale IMF
12	- Apply LSTM
13	- Train the model and generate forecast
14	- Inverse scale and store predictions
15	For each VMD mode:
16	- Scale mode
17	- Apply GRU
18	- Train the model and generate forecast
19	- Inverse scale and store predictions
20	**Step 5**: Combine Predictions
21	- Sum predictions across all IMFs and VMD modes
22	- Generate final forecasted signal $\widehat{y}$_{T+1}
23	**Step 6**: Evaluation
24	- Compare $\widehat{y}$_{T+1} with actual values

Appendix C

The pseudocode for the Decentralized LSTM-GRU Based Day-Ahead Price Forecasting Framework is given in Algorithm 3.

Algorithm 3. The pseudocode for the Decentralized LSTM-GRU Based Day-Ahead Price Forecasting Framework

	Input: Forecasted feature values obtained from Algorithm 2: X = [Load forecast (T+1), Dummy DoW (T+1), Generation features (T+1), Amount of Electricity Exchange (T+1), Previous day Spot Prices (T)] - Hyperparameters for deep learning models
	Output: Forecasted 24-h ahead LMP values: $\hat{y}$_{T+1} = [$\hat{p}$_1, $\hat{p}$_2, …, $\hat{p}$_{24}]
1	**Step 1**: Data Preprocessing
2	- Handle missing values and ensure numerical format
3	- Normalize the data using MinMaxScaler
4	**Step 2**: Decentralized Input Structure
5	For each feature in X:
6	- Reshape as separate input
7	**Step 3**: Model Architecture
8	For each feature input:
9	- Apply LSTM Layer → Dropout → GRU Layer
10	- Store outputs
11	- Concatenate all outputs
12	- Apply Dense Layer(s)
13	- Final Dense Layer → Output 24 values (Spot Prices for T+1)
14	**Step 4**: Model Training
15	- Compile and train model
16	**Step 5**: Prediction and Inverse Scaling
17	- Predict Spot Prices values on test set
18	- Inverse transform predictions to original scale
19	**Step 6**: Evaluation
20	- Compare $\hat{y}$_{T+1} with actual values

Data availability

Data will be made available on request.

References

Abdellatif, H., Hussein, A.E., Alawami, A.T., Abido, M.A., 2023. Real-time electricity market price prediction using improved ARIMA model. In: 2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe). IEEE, pp. 1–6. https://doi.org/10.1109/EEEIC/ICPSEurope57605.2023.10194648.

Ceni, A., 2025. Random orthogonal additive filters: a solution to the vanishing/exploding gradient of deep neural networks. IEEE Transact. Neural Networks Learn. Syst. https://doi.org/10.1109/TNNLS.2025.3538924.

Chen, K., Zhou, X.C., Fang, J.Q., Qin, L., 2017. Study on frequency characteristics of rotor systems for fault detection using variational mode decomposition. Int. J. Rotating Mach. 2017 (1), 7218646. https://doi.org/10.1155/2017/7218646.

Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y., 2014. On the properties of neural machine translation: encoder-decoder approaches. arXiv. https://doi.org/10.48550/arXiv.1409.1259 preprint arXiv:1409.1259.

da Silva Leite, A.L., de Lima, M.V.A., 2023. A garch model to understand the volatility of the electricity spot price in Brazil. Int. J. Energy Econ. Pol. 13 (5), 332–338. https://doi.org/10.32479/ijeep.14226.

Dai, Y., Yu, W., 2024. Short-term power load forecasting based on Seq2Seq model integrating Bayesian optimization, temporal convolutional network and attention. Appl. Soft Comput. 166, 112248. https://doi.org/10.1016/j.asoc.2024.112248.

de Castilho Braz, D.D., dos Santos, M.R., de Paula, M.B.S., da Silva Filho, D., Guarnier, E., Alípio, L.P., et al., 2024. Multi-source data ensemble for energy price trend forecasting. Eng. Appl. Artif. Intell. 133, 108125. https://doi.org/10.1016/j.engappai.2024.108125.

Diebold, F.X., Mariano, R.S., 1995. Com paring predictive accu racy. J. Bus. Econ. Stat. 13 (3), 253–263. https://doi.org/10.2307/1392185.

Dragomiretskiy, K., Zosso, D., 2013. Variational mode decomposition. IEEE Trans. Signal Process. 62 (3), 531–544. https://doi.org/10.1109/TSP.2013.2288675.

Eleftheriadis, P., Giazitzis, S., Kowal, J., Leva, S., Ogliari, E., 2024. Joint state of charge and state of health estimation using bidirectional lstm and bayesian hyperparameter optimization. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3410675.

Energidataservice web page. www.energidataservice.dk. (Accessed 20 November 2024).

Gong, H., Xing, H., Wang, Q., 2025. Enhanced forecasting method for realized volatility of energy futures prices: a secondary decomposition-based deep learning model. Eng. Appl. Artif. Intell. 146, 110321. https://doi.org/10.1016/j.engappai.2025.110321.

Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J., 2016. LSTM: a search space odyssey. IEEE Transact. Neural Networks Learn. Syst. 28 (10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924.

Hanifi, S., Cammarono, A., Zare-Behtash, H., 2024. Advanced hyperparameter optimization of deep learning models for wind power prediction. Renew. Energy 221, 119700. https://doi.org/10.1016/j.renene.2023.119700.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Math. Phys. Eng. Sci. 454, 903–995. https://doi.org/10.1098/rspa.1998.0193, 1971.

Huang, C.J., Shen, Y., Chen, Y.H., Chen, H.C., 2021. A novel hybrid deep neural network model for short-term electricity price forecasting. Int. J. Energy Res. 45 (2), 2511–2532. https://doi.org/10.1002/er.5945.

Janczura, J., Puć, A., 2023. ARX-GARCH probabilistic price forecasts for diversification of trade in electricity markets—variance stabilizing transformation and financial risk-minimizing portfolio allocation. Energies 16 (2), 807. https://doi.org/10.3390/en16020807.

Jayanth, T., Manimaran, A., 2024. Developing a novel hybrid model double exponential smoothing and dual attention encoder-decoder based Bi-Directional gated recurrent unit enhanced with bayesian optimization to forecast stock price. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3435683.

Kaya, M., Karan, M.B., Telatar, E., 2023. Electricity price estimation using deep learning approaches: an empirical study on Turkish markets in normal and Covid-19 periods. Expert Syst. Appl. 224, 120026. https://doi.org/10.1016/j.eswa.2023.120026.

Khan, Z.A., Fareed, S., Anwar, M., Naeem, A., Gul, H., Arif, A., Javaid, N., 2020. Short term electricity price forecasting through convolutional neural network (cnn). In web, artificial intelligence and network applications. Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020). Springer International Publishing, pp. 1181–1188.

Khan, S., Aslam, S., Mustafa, I., Aslam, S., 2021. Short-term electricity price forecasting by employing ensemble empirical mode decomposition and extreme learning machine. Forecasting 3 (3), 28. https://doi.org/10.3390/forecast3030028.

McHugh, C., Coleman, S., Kerr, D., 2022. Hourly electricity price forecasting with NARMAX. Machine Learning with Applications 9, 100383. https://doi.org/10.1016/j.mlwa.2022.100383.

Moradzadeh, A., Mouhammadpourfard, M., Weng, Y., Pol, S., Muyeen, S.M., 2025. Hybrid deep learning model for accurate short-term electricity price forecasting. In: 2025 IEEE Texas Power and Energy Conference (TPEC). IEEE, pp. 1–6. https://doi.org/10.1109/TPEC63981.2025.10906930.

Osório, G.J., Lotfi, M., Shafie-Khah, M., Campos, V.M., Catalão, J.P., 2018. Hybrid forecasting model for short-term electricity market prices with renewable integration. Sustainability 11 (1), 57. https://doi.org/10.3390/su11010057.

Panapakidis, I.P., Dagoumas, A.S., 2016. Day-ahead electricity price forecasting via the application of artificial neural network based models. Appl. Energy 172, 132–151. https://doi.org/10.1016/j.apenergy.2016.03.089.

Plakas, K., Karampinis, I., Alefragis, P., Birbas, A., Birbas, M., Papalexopoulos, A., 2023. A predictive fuzzy logic model for forecasting electricity day-ahead market prices for scheduling industrial applications. Energies 16 (10), 4085. https://doi.org/10.3390/en16104085.

Pourdaryaei, A., Mohammadi, M., Mubarak, H., Abdellatif, A., Karimi, M., Gryazina, E., Terzija, V., 2024. A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market. Expert Syst. Appl. 235, 121207. https://doi.org/10.1016/j.eswa.2023.121207.

Qi, C., Ren, J., Su, J., 2023. GRU neural network based on CEEMDAN–wavelet for stock price prediction. Appl. Sci. 13 (12), 7104. https://doi.org/10.3390/app13127104.

Rajan, P., Chandrakala, K.V., 2021. Statistical model approach of electricity price forecasting for indian electricity market. In: 2021 IEEE Madras Section Conference (MASCON). IEEE, pp. 1–5. https://doi.org/10.1109/MASCON51689.2021.9563474.

Setayesh Nazar, M., Eslami Fard, A., 2021. The adaptive neuro-fuzzy inference system model for short-term load, price, and topology forecasting of distribution system. Application of Machine Learning and Deep Learning Methods to Power System Problems, pp. 321–343.

Sola, J., Sevilla, J., 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. IEEE Trans. Nucl. Sci. 44 (3), 1464–1468. https://doi.org/10.1109/23.589532.

Taheri, S., Talebjedi, B., Laukkanen, T., 2021. Electricity demand time series forecasting based on empirical mode decomposition and long short-term memory. Energy Eng. J. Assoc. Energy Eng.: Journal of the Association of Energy Engineering 118 (6), 1577–1594. https://doi.org/10.32604/EE.2021.017795.

Tan, Y.Q., Shen, Y.X., Yu, X.Y., Lu, X., 2023. Day-ahead electricity price forecasting employing a novel hybrid frame of deep learning methods: a case study in NSW, Australia. Elec. Power Syst. Res. 220, 109300. https://doi.org/10.1016/j.epsr.2023.109300.

Wang, K., Yu, M., Niu, D., Liang, Y., Peng, S., Xu, X., 2023. Short-term electricity price forecasting based on similarity day screening, two-layer decomposition technique and Bi-LSTM neural network. Appl. Soft Comput. 136, 110018. https://doi.org/10.1016/j.asoc.2023.110018.

Waqas, M., Humphries, U.W., 2024. A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX, 102946. https://doi.org/10.1016/j.mex.2024.102946.

Xiong, X., Qing, G., 2023. A hybrid day-ahead electricity price forecasting framework based on time series. Energy 264, 126099. https://doi.org/10.1016/j.energy.2022.126099.

Xu, Y., Huang, X., Gao, Z., Mohamed, M.A., Jin, T., 2025. A novel electricity price forecasting approach based on multi-attention feature fusion model optimized by variational mode decomposition. Measurement, 117596. https://doi.org/10.1016/j.measurement.2025.117596.

Yang, H., Schell, K.R., 2020. HFNet: forecasting real-time electricity price via novel GRU architectures. In: 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, pp. 1–6. https://doi.org/10.1109/PMAPS47429.2020.9183697.

Yang, W., Chen, Z., Zhao, H., Chen, S., Shi, C., 2025. Feature fusion method for rock mass classification prediction and interpretable analysis based on TBM operating and cutter wear data. Tunn. Undergr. Space Technol. 157, 106351. https://doi.org/10.1016/j.tust.2024.106351.

Zhang, Y., Deng, C., Zhao, R., 2020. A novel integrated price and load forecasting method in smart grid environment based on multi-level structure. Eng. Appl. Artif. Intell. 95, 103852. https://doi.org/10.1016/j.engappai.2020.103852.

Zhang, H., Hu, W., Cao, D., Huang, Q., Chen, Z., Blaabjerg, F., 2021. A temporal convolutional network based hybrid model of short-term electricity price forecasting. CSEE Journal of Power and Energy Systems. https://doi.org/10.17775/CSEEJPES.2020.04810.