An effective dimensionality reduction approach for short-term load forecasting

Yang, Yang, Wang, Zijin, Gao, Yuchao, Wu, Jinran, Zhao, Shangrui and Ding, Zhe

Yang, Y., Wang, Z., Gao, Y., Wu, J., Zhao, S. and Ding, Z. (2022). An effective dimensionality reduction approach for short-term load forecasting. *Electric Power Systems Research, 210*, Article 108150. https://doi.org/10.1016/j.epsr.2022.108150

An effective dimensionality reduction approach for short-term load forecasting

Yang Yang[a], Zijin Wang[a], Yuchao Gao[a], Jinran Wu[b,*], Shangrui Zhao[c], Zhe Ding[d]

[a] *College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, P. R. China*
[b] *School of Mathematical Sciences, Queensland University of Technology, Brisbane 4001, Australia*
[c] *School of Science, Wuhan University of Technology, Wuhan 430070, P. R. China*
[d] *School of Computer Science, Queensland University of Technology, Brisbane 4001, Australia*

Abstract

Accurate power load forecasting has a significant effect on a smart grid by ensuring effective supply and dispatching of power. However, electric load data generally possesses the characteristics of nonlinearity, periodicity, and seasonality. In particular, for complex electric load systems, the presence of redundant information potentially reduces the real pattern extraction for load forecasting. Bearing in mind these issues, we propose an effective forecasting model in which a feature extraction module is introduced that is combined with the variational mode decomposition (VMD) with the variational autoencoder (VAE). In this combination, VMD is utilized for decomposing complex load series and VAE is used to filter the redundant information (noises) from each decomposed series. With two real data sets from China, we demonstrate that the proposed model can achieve highly accurate predictions, as we find the mean absolute percentage error (MAPE) values for one-step-ahead prediction to be 1% (Nanjing) and 0.8% (Taixing), respectively.

Keywords: Deep learning, Decomposition-ensemble method, Feature extraction, Load forecasting

1. Introduction

In recent years, given that there are insufficient energy resources, the establishment of reliable energy management system (EMS) has become the focus [1, 2, 3], and high-performance power load forecasting is the basis of EMS automatic management. Power load forecasting is the most important issue in EMS [4, 5, 6] and it affects power generation, power dispatch, power purchase, resource management, grid security, and stability [7]. Therefore, accurate electric load prediction can ensure the safety and reliability of EMS. In addition, the improvement in the accuracy of power load forecasting can further reduce the cost of power operations. According to Xiao et al. [8] and Dong et al. [9], for every 1% increase in the accuracy of power forecasts, the cost of smart grid operations will likely be reduced by millions of

*Corresponding author
Email addresses: yyang@njupt.edu.cn (Yang Yang), 1321058617@njupt.edu.cn (Zijin Wang), b18040933@njupt.edu.cn (Yuchao Gao), j73.wu@qut.edu.au (Jinran Wu), zhaosr@whut.edu.cn (Shangrui Zhao), zhe.ding@hdr.qut.edu.au (Zhe Ding)

dollars. In summary, a power load forecasting model with excellent performance is essential for building a reliable and efficient EMS [10].

1.1. Literature reviews

According to the duration of forecast, power load forecasting can be divided into long-term load forecasting (LTLF) [11], medium-term load forecasting (MTLF) [12], and short-term load forecasting (STLF) [13, 14, 15]. STLF serves to formulate power generation plans and transmission schemes. It is an important part of EMS. This paper focuses on SLTF **based on time series modelling** due to its immense practicality. Substantial research has been conducted by several researchers to advance to forecasting performance of the power load forecasting model [16, 17, 18].

Since power load data often exhibit complex characteristics—non-stationary, non-linearity, and multi-seasonality—power load data has numerous dimensions. Before forecasting, it is necessary to preprocess electric load data, as this is expected to strengthen prediction performance. The traditional power load data processing methods include modal decomposition and data dimensionality reduction. Decomposition algorithms such as empirical mode decomposition (EMD) [19], ensemble empirical mode decomposition (EEMD) [20], and complete ensemble empirical mode decomposition (CEEMD) [21] can handle nonlinear and unstable sequences well and have been extensively utilized in time series prediction since they were proposed [20].

Principal component analysis (PCA) [22] and linear discriminant analysis (LDA) [23] are traditional unsupervised and supervised linear dimensionality reduction methods, respectively. As a traditional generation model, variational autoencoder (VAE) proposed by Kingma and Welling [24] has received much attention in recent years for its dimensionality reduction ability. Compared with autoencoder (AE), VAE adds a few restrictions in the encoding process, thereby forcing the generated latent vector to roughly follow a standard normal distribution [25, 26]. Moreover, VAE retains the features that are important for prediction and eliminates noise. Xu et al. [27] compared the hidden layer learning of hidden layer in conventional AE and VAE. VAE can generate different data to strengthen the sample data that shows better performance. Laubscher and Rousseau [28] combined VAE with deep neural networks (DNNs) to forecast, where VAE is responsible for generating low-dimensional encoded data. In these studies, VAE reduces the dimensionality of input data and compresses high-dimensional data to extract features.

Power load forecasting models are generally divided into two main classifications: traditional statistical methods [29] and deep learning methods [30]. Statistical methods include regression prediction, trend extrapolation, and time-series analysis methods, in which the method of analysing time series includes auto regression (AR) [31] and the auto-regressive integrated moving average model (ARIMA) [32]. ARIMA is constituted of AR and the moving average model (MA). The model coefficients can be updated according to least squares or gradient descent. ArunKumar et al. [33] used ARIMA to forecast epidemiological trends and confirmed the existence of seasonality in data. Abdalla et al. [34] used ARIMA

to analyze the short-term uncertainty that existed in wind forecasts, and various methods are proposed for its influence. However, for complex nonlinear systems, models based on traditional statistical methods are not entirely applicable, and the forecast performance is poor [35, 36].

Since the 1980s, numerous scholars have begun to study deep learning methods and successfully apply them to power load forecasting, including support vector machines (SVM) [37, 38] and neural networks [39, 40]. Extreme learning machine (ELM) is an emerging generalized single hidden layer feed-forward neural network learning algorithm, which can generate random hidden variable parameters to calculate output weights and is widely used in forecasting [41]. Yang et al. [42] improved the whale optimization algorithm and applied it with the robust ELM. Further, Yang et al. [43] implemented general robust parameterized ELM by redesigning the statistical framework in robust parameterization to improve the forecasting effect of ELM. Wu et al. [44] proposed a new SVR, which incorporates a linear cost function and the insensitivity parameter.

In addition to these above methods, long short-term memory network (LSTM) is also utilized for power load prediction. LSTM is a time loop neural network. During the training process of long sequences, LSTM can provide solutions for gradient disappearance and gradient explosion [45]. To improve prediction accuracy, numerous scholars have used modal decomposition algorithms to improve LSTM. Cheng et al. [46] combined EMD and backpropagation long short-term memory (B-LSTM) to develop estimation and prediction models. Jana et al. [47] proposed a granular deep learning approach consisting of maximal overlap discrete wavelet transformation (MODWT) and LSTM for predicting the energy consumption. However, when the data is noisy, using these traditional decomposition algorithms will likely lead to lead to modal aliasing.

Variational modal decomposition (VMD) is widely used due to its strong robustness to data and its ability to avoid modal aliasing [48]. In recent years, scholars have been concerned about improving the feature extraction efficiency of VMD. Li et al. [49] proposed a two-stage short-term wind power forecast method by optimizing VMD parameters through the flower pollination algorithm (FPA). Zhang et al. [50] introduced a two-layer decomposition technique based on the combination of VMD and ensemble empirical modal decomposition (EEMD). Therefore, it is effective to combine VAE with VMD to improve the ability to extract features.

In summary, the current demand for accurate forecasting of power loads is increasing. First, more advanced deep learning methods are required to build hybrid models to improve prediction accuracy. Second, the extraction of more effective features from data to meet the forecast demand of different regions is another problem.

1.2. Contributions of this Study

In response to the above problems, a hybrid load forecasting model called VMD-VAE-LSTM is proposed in this paper. In this model, VMD as the data processing module splits data into multiple subsequences, VAE performs feature extraction on the decomposed sequence, and finally LSTM is used for

prediction. The results reveal that the hybrid model we proposed can effectively improve the forecasting accuracy. This paper makes the following contributions to literature:

(a) To improve the accuracy of electric load forecasting, a new hybrid model is proposed in this paper. The model combines LSTM with VMD and VAE. The raw electric load data is decomposed utilizing VMD, and the features of the input sequence are extracted using VAE, which effectively improves the prediction accuracy of LSTM.

(b) The introduction of VAE into the decomposition sequence to extract effective features greatly reduces the dimension of the data, particularly for the extraction of efficacious information from components.

(c) The proposed hybrid load forecasting model is used to execute multi-step-ahead forecasting on two real power load data sets in Nanjing and Taixing. Compared with AE, VAE significantly improved the prediction accuracy in both examples. In the one, three, and five-step-ahead forecasting of Example 1, MAPE decreased by 61.5%, 23.8%, and 32.1%, respectively. In the one, three, and five-step-ahead forecasting of Example 2, MAPE decreased by 55.6%, 40.9%, and 22.2%, respectively.

1.3. The structure of the paper

The remainder of this paper is organized in the following manner: Section 2 details the methodology employed in this study. Section 3 presents the specific process of VMD-VAE-LSTM. Section 4 presents the application and the experiment error indexes of the proposed model on the electric load data of Nanjing and Taixing. Finally, Section 5 summarizes the above experiments and provides criticism, discussion, and suggestions.

2. Methodology

2.1. Variational mode decomposition

EMD has been widely applied in nonlinear and nonstationary signal analysis, but the intrinsic mode function(IMF) components obtained by this method have mode aliasing. The appearance of mode aliasing leads to false time-frequency distribution. In addition, the IMF with features cannot be accurately generated. Considering this, Dragomiretskiy and Zosso [51] proposed a VMD method that can transfer the acquisition process of signal components to the variational framework. Simultaneously, the method is a completely no-recursive variational pattern decomposition model and a simultaneous extraction mode. VMD reproduces the input signal by searching the component and centre frequency. After demodulation to the baseband, each mode is smooth. Compared with other decomposition algorithms, VMD has better performance in terms of the accuracy of decomposing complex data and the anti-interference ability of noise.

In this paper, the electric load data $L(t)$ is decomposed by VMD into sub-sequences (or modes), where u_m represents the mth mode after decomposition, m belongs to $[1, n]$, and w_m represents the mth centre frequency. The detailed realization steps of VMD method are described below.

Step 1: VMD constructs a variational problem. K mode components are generated by decomposing the raw signal f. Each component should have a limited bandwidth of the central frequency, and the sum of these estimated bandwidths should be the smallest. Therefore, the constraint condition that the sum of all mode components is equal to the raw signal can be assumed,

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left\{ [\nabla(t) + \frac{j}{t\pi}] * u_k(t) \right\} e^{-j\omega_k t} \right\|_2^2 \right\}, \\ &s.t. \sum_{k=1}^{K} u_k = f, \end{aligned} \tag{1}$$

where $\{u_k\}$ and $\{\omega_k\}$ correspond to the kth model component and centre frequency, respectively, K is the number of decomposed modes, ∇ denotes the Dirac distribution, and $*$ is a convolution operator.

Step 2: By introducing the Lagrangian operator λ, the constrained variational problem is transformed into an unconstrained variational problem, and the augmented Lagrangian expression is as given below:

$$\begin{aligned} &L(\{u_k\}, \{\omega_k\}, \lambda) = \\ &\alpha \left\{ \sum_k \left\| \partial_t [(\nabla(t) + \frac{j}{t\pi}) * u_k(t)] e^{-j\omega_k t} \right\|_2^2 \right\} + \\ &\|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle, \end{aligned} \tag{2}$$

where α is the quadratic penalty factor for reducing Gaussian noise interference.

Step 3: VMD optimizes each modal component and the center frequency, by the Fourier isometric transform and the alternate direction method of multipliers (ADMM). By searching the saddle point of the augmented Lagrangian function, the u_k, ω_k, and λ are optimized alternately after iteration. The specific equation is shown below:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}, \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}, \tag{4}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)), \tag{5}$$

where τ is the noise tolerance that meets the fidelity requirements after signal decomposition, $\hat{u}_k^{n+1}(\omega)$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}(\omega)$ correspond to $u_k^{n+1}(t)$, $u_i(t)$, $f(t)$, and $\lambda(t)$ in Fourier transform, respectively.

2.2. Variational autoencoder

VAE is a special AE based on variable Bayesian reasoning [24]. Like AE, VAE is composed of an encoder and a decoder, which are two kinds of neural networks with different structures. The encoder is responsible for extracting feature information. The decoder reconstructs this information to achieve an outcome, which is similar to the original data.

The encoder represents $z = \sigma(\omega x + b)$, where x denotes the original data, and z denotes the extracted feature information. The decoder represents $x' = \sigma'(\omega' z + b')$, where x' denotes the reconstruction result. VAE learns the feature through the loss function $Loss(x, x') = \|x - x'\|^2$.

Different from AE, VAE has the characteristics of probabilistic AE and generative AE. Therefore, VAE is divided into a corresponding reasoning network and a generation network. The encoder is an inference network. The hidden variables are obtained by the variational inference treatments on original data, and the variational probability distribution is further obtained from this data. The decoder is called the generation network, which generates data approximate to original data.

In addition, VAE has better robustness than AE. This is because Gaussian noise is added to the encoder of VAE, which enhances its ability to resist noise in an actual scenario.

2.3. Long short-term memory network

The LSTM network is developed depending on the recurrent neural network (RNN), which controls the transmission state through the gate structure [52]. Compared with ordinary neural networks, RNN can effectively manage sequential data. RNN has the advantage of processing data that varies in terms of sequence, which can take advantage of the past timestamp to estimate current load requirements. However, RNN has disadvantages in that it has the potential situation of gradient disappearance and gradient explosion. These issues are improved by the introduction of gates in the structure of the LSTM. The LSTM is implemented in three steps, as described below.

Step 1: The forget gate ($f_g = \sigma(W_f[h_{g-1}, x_g] + b_f)$) decides which information will be discarded from the cell state. The forget gate determines the retention of the previous cell state C_{g-1} by learning the output h_{g-1} of the previous reservoir and the input x_g of the current sequence. The output result determines the dependence of the current cell state on C_{g-1}. 0 is abandoned and 1 is completely passed.

Step 2: The preparation of the update of the cell state is done in two parts. First, the input gate ($i_g = \sigma(b_i + W_i[h_{g-1}, x_g])$) determines which values to update and prepares a candidate vector ($\tilde{C}_g = \tanh(b_C + W_C[h_{g-1}, x_g])$). Second, forget gate and C_{g-1}, input gate and $\tilde{C}_{g-1}$ are multiplied and summed to obtain updated results ($C_g = f_g \times C_{g-1} + i_g \times \tilde{C}_g$).

Step 3: The output gate ($o_g = \sigma(W_o[h_{g-1}, x_g] + b_o)$) determines which parts of the cell state to be output of current state. The final hidden layer output ($h_g = o_g \times tanh(C_g)$) is obtained by multiplying o_g with C_g, and it acts on the next cell state.

3. The proposed model

Electric load forecasting uses historical electric load data to estimate the load data at a certain time in the future. The result of STLF spans a few minutes to a week. STLF is widely employed in daily load scheduling in various fields due to its strong practical value. As mentioned above, there are two difficulties in STLF. One is the development of a model that can satisfy the various complex nonlinear

load data, and the other is the forecasting performance of the existing models to be improved. Because of these difficulties, LSTM is designed to predict mass data. In addition, VMD is used to extract feature sub-sequences, and VAE is adopted to reduce the dimensionality of data. Finally, VMD-VAE-LSTM is proposed. The specific process of the corresponding framework establishment can be summarized in five steps as described below.

Step 1: Feature extraction. The K value was set manually, which is mentioned in Section 2.1. Then, the desired sub-sequence set can be achieved from VMD. VMD is employed to extract the sub-sequences with features from power load data $X_p(p = 1, 2, 3, \cdots, S)$ according to frequency, where S is the total amount of raw data. Finally, VMD outputs N sub-sequences Mod_j $(j = 1, 2, 3, \cdots, N)$.

Step 2: Dimensionality reduction. According to the proportion of 80%-20%, the training set and testing set is generated by the division of each sub-sequence Mod_j $(j = 1, 2, 3, \cdots, N)$ in Step 1. Then, the training set is fed into VAE for the establishment of the network, and the hyper-parameters of VAE are adjusted by the fitting degree of the testing set on VAE. Finally, VAE outputs the reconstruction results R_r $(r = 1, 2, 3, \cdots, T)$ based on the best reconstruction outcomes and the least dimension.

Step 3: Model training. According to the segmentation data set method in Step 2, the reconstruction results in Step 2 are divided into the training set and testing set. To achieve different prediction objectives, the training set and testing set are divided into sample set and label set, respectively. LSTM was trained and verified by the training set and the testing set separately. Then, the hidden layer parameters are adjusted adaptively by calculating the loss function of prediction results and label set in each iteration. Finally, the training of the prediction model is conducted.

Step 4: Multi-step forecasting. The prediction result Per_l $(l = 1, 2, 3, \cdots, M)$ is obtained by inputting each sub-sequence with its testing set. Multi-step forecasting can be finally realized according to the label sets with different prediction objectives.

Step 5: Accumulation results. The output of each sub-sequence forecasting Per_l are added to calculate the forecasting result of raw data.

Remark 1. It should be noted that the choice of the parameter K in Step 1 (Feature extraction) would determine the performance of our proposed dimensionality reduction approach. In our proposed procedure, the tuning value of K is set by the empirical method. Some advanced methods like successive approximation [53] can be used to optimize the parameter K for further improving the effectiveness of our proposed approach.

For a more clear and comprehensive expression, the specific steps of VMD-VAE-LSTM are displayed more intuitively and concisely. Figure 1 illustrates the specific steps of the developed model. Figure 1 shows the specific steps of the developed model.

The network structure of VMD-VAE-LSTM is illustrated in Figure 2. The specific implementation process of VMD-VAE-LSTM is provided in Algorithm 1. In this paper, the value of K is used in accordance

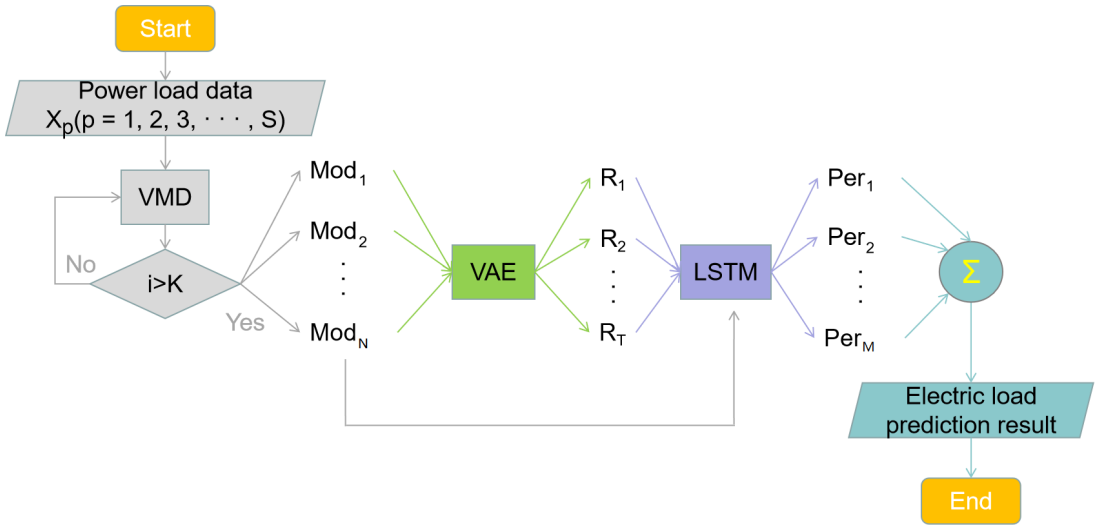


Figure 1: The flowchart of the VMD-VAE-LSTM model.

with the empirical method, which is determined by the experimental results through several experiments.

To verify the applicability and reliability of this method, the regional load data sets of Nanjing and Taixing are used. Section 4 provides an in-depth explanation of the above methods and forecasting results.

4. Case study

In this section, the experimental results are comprehensively analyzed by answering the following questions.

(1) How do the different decomposition algorithms affect load forecasting? Can VMD provide higher forecasting accuracy?

(2) How does the reduction of data dimensionality influence load forecasting? Does VAE perform better in load forecasting?

(3) Compared with other popular forecasting models, how does VMD-VAE-LSTM perform?

By designing comparative experiments, the performance of VMD-VAE-LSTM is further verified. All models involved in this investigation are listed in Table 1. These models can be divided into three categories: benchmark models, popular models, and the model proposed in this paper. The detailed hyperparametric settings of different benchmark models are provided in Appendix A. Further, the adopted data are from Nanjing and Taixing, respectively. The overall verification process is conducted using Matlab R2020a and PyCharm Community Edition 2021.1.3 x64 environment with Windows 10 and a 2.30 GHz Intel Core i5-8300H CPU, with 64-bit support and 8GB RAM.

To objectively measure the simulation effect and fitting degree of different models, various indicators are adopted in Appendix B to quantitatively analyze the forecasting results. The mathematical expressions of the error indexes are described in Appendix B, where S represents the statistics of data

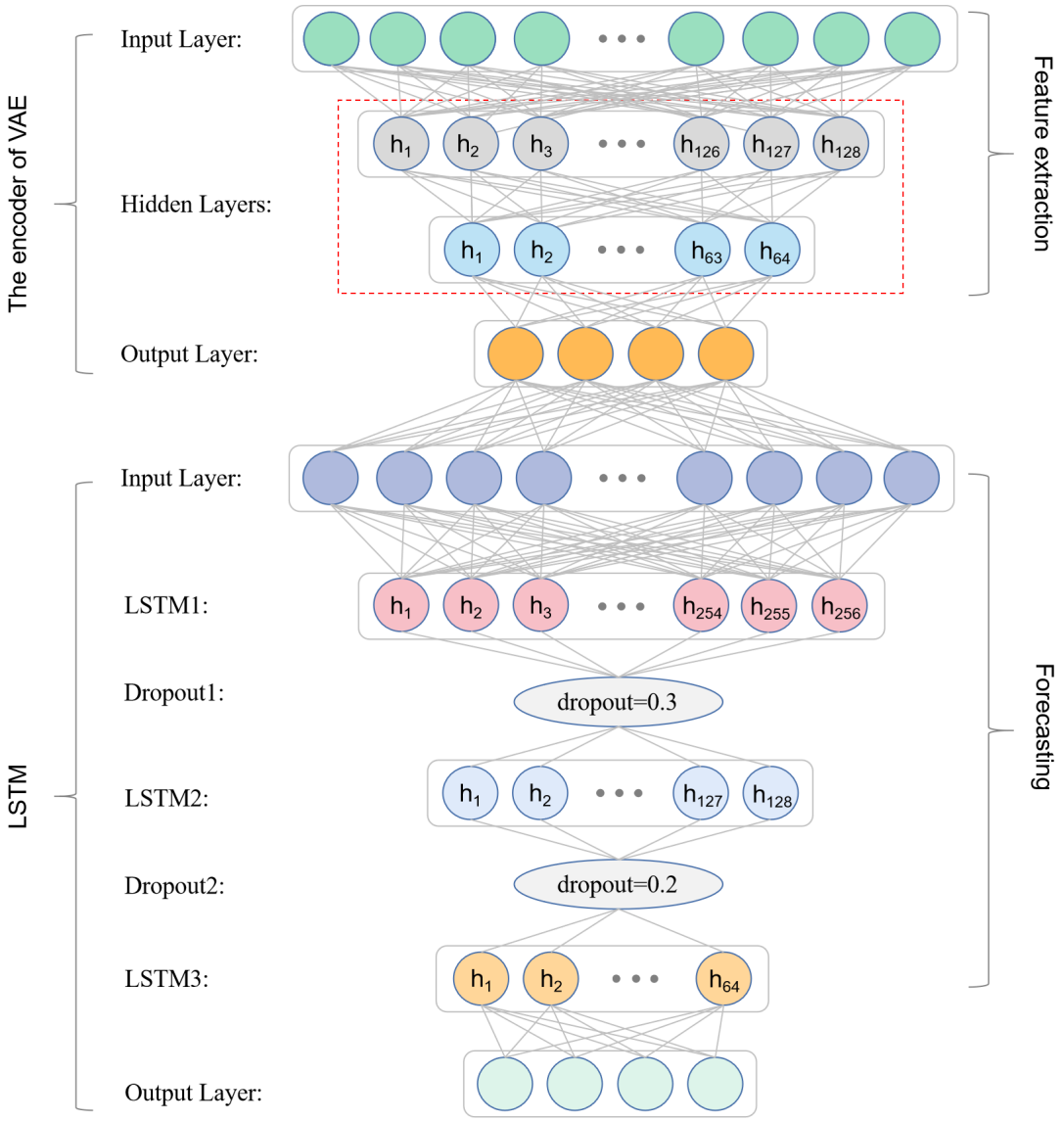


Figure 2: The network structure of the VMD-VAE-LSTM model.

quantity, P and $\hat{P}$ indicate the true and forecasted values, respectively. Specifically, when root mean square error (RMSE) or mean absolute error (MAE) gradually approaches 0, the prediction result is also close to the actual result. In addition, MAPE indicates the quality of the model in terms of percentage. When the value of MAPE approaches or exceeds 100%, it indicates the model has poor performance. The forecasting accuracy of the model increases with the decrease in MAPE.

Note that, that popular EEMD, CEEMD, and CEEMDAN are designed to show the advantage of VMD. The performance verification of VMD-VAE-LSTM is studied with respect to two common models: RNN and LSTM, where the learning rate is 0.01, the iteration round is set to 200, and hidden neurons are $(256, 128, 64, 1)$. The goal of this research is to utilize historical electric load data to forecast the electric load data before the one-step, three-step, and five-step-ahead forecasting, respectively.

Algorithm 1 The VMD-VAE-LSTM model

Input: Power load data from Nanjing or Taixing

Output: Predicted value of power load data

if $D_P \neq 1$ **then**
  Convert P to one-dimensional vector
end if
Set the value of expected patterns (K)
Obtain K decomposed sub-sequences (S) with $L \times 1$, L is the length of P
for all $i = 1, 2, 3, \cdots, K$ **do**
  Set the size (W) of window in Sliding Window Algorithm
  Obtain V matrix with $(L - W) \times W$
  Set the desired dimension (R)
  Obtain the result of dimension reduction matrix ($\hat{V}$) with $(L - W) \times R$
end for
for all $i = 1, 2, 3, \cdots, K$ **do**
  Segment $\hat{V}$ into samples (A) and labels (B)
  Segment (A) and (B) into training set(A_a,B_a) and testing set(A_e,B_e), respectively
  Create batch data
  Training model with A_a and S
  Feed the trained model with A_e and S
  Obtain the predicted value
end for

The experimental outcomes of both data sets are analysed from three perspectives: the improvement of prediction accuracy from different decomposition algorithms, the dimensions of data, and the forecasting accuracy of various popular forecasting models. The subsections below describe these aspects.

4.1. Example 1: Nanjing

Example 1 adopts the load data from the total electric consumption in Nanjing. The electric load data in Nanjing is collected every 15 minutes. The Nanjing data set has 22272 data points (from 00:00 on November 1, 2002 to 23:45 on June 26, 2003). The descriptive statistics of the data set in Nanjing is presented in Table 2. This data set is split 80% and 20% to produce two subsets. One is the training set with 17862 data points, and the other is the testing set with 4410 data points. Table 3 presents the error indexes of the experiment in Example 1. The actual power load data and prediction results based on VMD-VAE-LSTM from Nanjing are presents in Figure 4. Then, the experiment is specifically analyzed.

Table 1: All models adopted in the experiment.

Model	Abbreviation	Definition
Benchmark Models	RNN	Recurrent Neural Network [54]
	LSTM	Long Short-Term Memory Networks [52]
Popular Models	EMD-LSTM	LSTM with Empirical Mode Decomposition [46]
	EEMD-LSTM	LSTM with Ensemble Empirical Mode Decomposition [20]
	CEEMD-LSTM	LSTM with Complete Ensemble Empirical Mode Decomposition [21]
	CEEMDAN-LSTM	LSTM with Complete Ensemble Empirical Mode Decomposition with Adaptive Noise [55]
	VMD-LSTM	LSTM with Variational Mode Decomposition [51]
	VMD-AE-LSTM	LSTM with Variational Mode Decomposition and Autoencoder [56]
Proposed Models	VMD-VAE-LSTM	LSTM with Variational Mode Decomposition and Variational Autoencoder

Table 2: The descriptive statistics of the Nanjing data set.

data set	Size	Min.	Max.	Median	Mean	Std. Dev.
Total data	22272	1154.501	3354.927	1910.669	1921.220	405.395
Training data	17862	1154.501	3354.927	1786.841	1849.930	395.040
Testing data	4410	1394.058	3221.731	2145.908	2209.973	306.547

The unit of load data in Nanjing data set is MW.

4.1.1. Comparison among EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM

To verify the improvement of VMD in the prediction performance of VMD-VAE-LSTM, several decomposition algorithms are designed for comparison in this experiment. In this experiment, the empirical method is used to test several setups of the K value (the number of expected patterns) in the decomposition algorithm. Figure 3 presents the decomposition effect of VMD. Because the original data set is too large, it is not easy to display clearly. 960 data points (from 00:00 on November 1, 2002 to 23:45 on November 4, 2002) are selected for display. As depicted in Figure 3, the original data set is decomposed into nine IMFs and one residual component. These components have periodic and regular components, which have advantages for forecasting. Figure 4 exhibits the power load forecasting curves in this comparison.

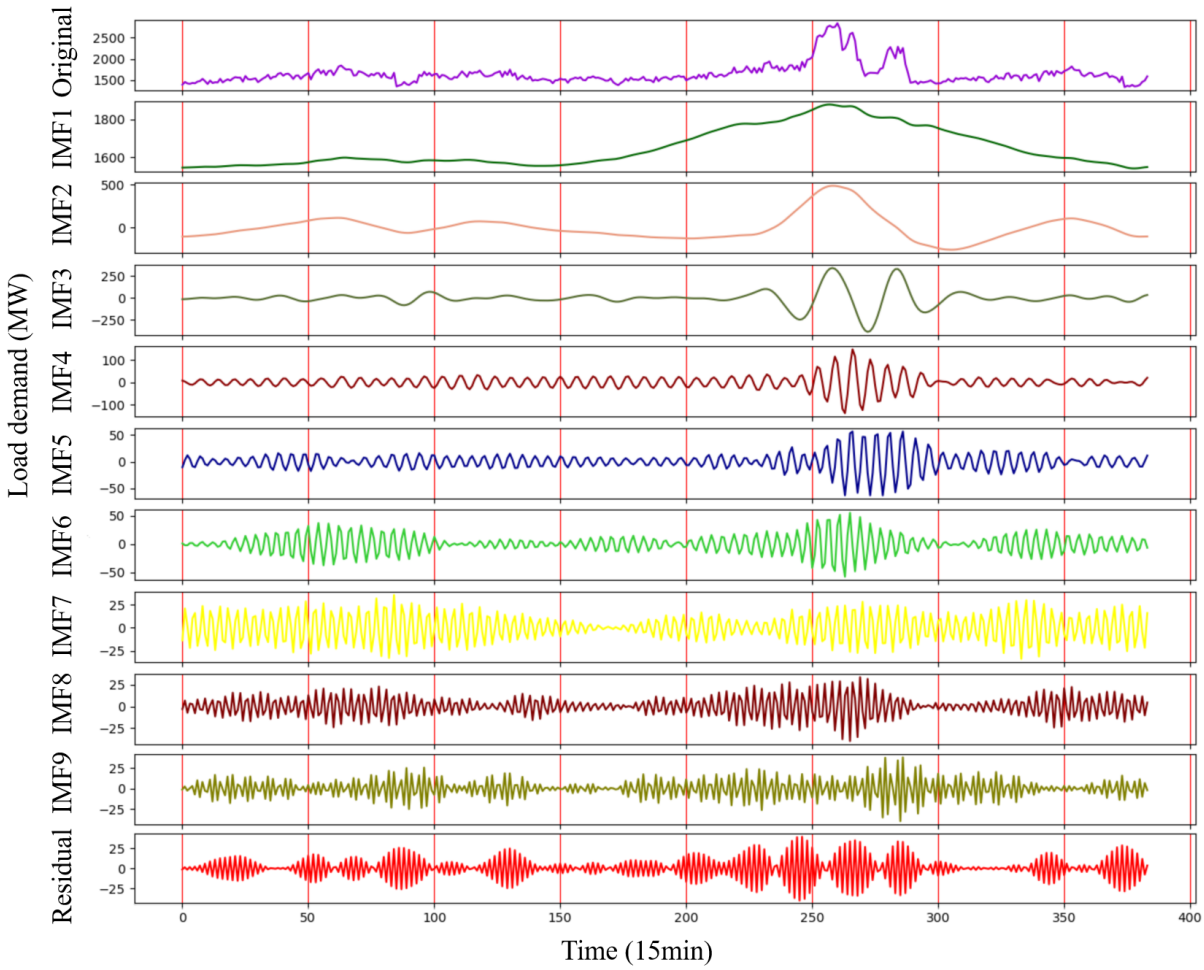
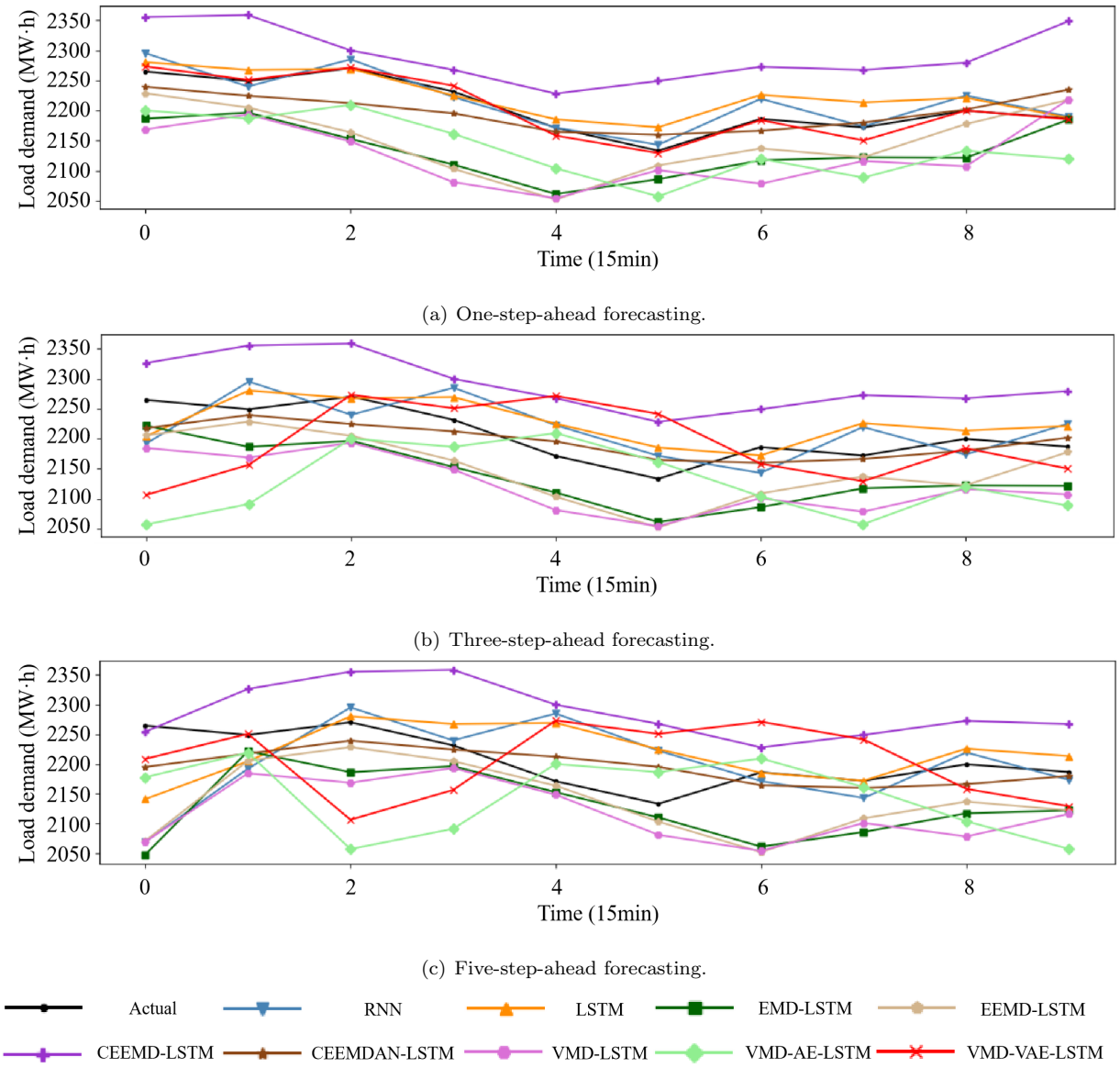


Figure 3: VMD components of the partial Nanjing data set (The time of the partial Nanjing data set is from 00:00 on November 1, 2002 to 23:45 on November 4, 2002).

The comparison reveals that the model with VMD has the best effect on prediction than other models with different decomposition algorithms. Compared with the model that only the decomposition algorithm, Figure 4 reveals that the electric load forecasting curve of VMD-LSTM is closer to the actual power data curve than the other four curves. In Table 3, VMD-LSTM obtains the smallest error indexes than other models only using decomposition algorithms. The goodness-of-fit (R^2) of VMD-LSTM in one-step, three-step, and five-step-ahead forecasting are 0.930, 0.928, and 0.914, respectively. The MAPE values are 0.032, 0.031, and 0.031, respectively. The MAE values are 69.638, 66.045, and 68.148, respectively. The RMSE values are 80.092, 81.327, and 88.832, respectively. The outcomes verify that LSTM with decomposition algorithm has a better prediction effect than other comparison models. Further, VMD can be used as a better data processing unit to obtain data that is more suitable for prediction.



(a) One-step-ahead forecasting.

(b) Three-step-ahead forecasting.

(c) Five-step-ahead forecasting.

Figure 4: One-, three- and five-step-ahead forecasting of various hybrid models in the Nanjing data set (The load data of Nanjing is from 23:15 on June 9, 2003 to 1:30 on June 10, 2003).

4.1.2. Comparison among VMD-LSTM, VMD-AE-LSTM, and VMD-VAE-LSTM

In order to examine the effectiveness of VAE in improving prediction performance, by reducing the dimension of data, VMD-VAE-LSTM is designed to compare with VMD-LSTM and VMD-AE-LSTM. It is evident from Figure 4 that VMD-AE-LSTM has a better fitting degree than VMD-LSTM and VMD-VAE-LSTM has better fitting degree than VMD-AE-LSTM. These indicate that the prediction effect of this model is better than that of other models. In Table 3, the R^2 of VMD-VAE-LSTM in one-step, three-step and five-step-ahead prediction are 0.993, 0.98, and 0.970, respectively. The MAPE values are 0.010, 0.016, and 0.019, respectively. The MAE values are 21.194, 35.032, and 41.152, respectively. The RMSE values are 25.747, 43.276, and 52.324, respectively. Compared with VMD-AE-LSTM, VMD-

VAE-LSTM significantly improved the prediction accuracy in both examples. In the one-step ahead prediction, MAPE decreased by 61.5%. In the three-step ahead prediction, MAPE decreased by 23.8%. In the five-step ahead prediction, MAPE decreased by 32.1%. In conclusion, VMD-VAE-LSTM is more accurate.

4.1.3. Comparison among EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, VMD-LSTM, and VMD-VAE-LSTM

In this section, further comparison is conducted between five different popular models and the developed model—EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, VMD-LSTM, and VMD-VAE-LSTM.

In contrast, the prediction result of VMD-VAE-LSTM is closer to the actual power load than other contrast models. For example, in Table 3, the R^2 of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM are (0.881, 0.763, 0.653), (0.908, 0.806, 0.843), (0.900, 0.835, 0.649), (0.918, 0.908, 0.892), and (0.930, 0.928, 0.914), (The values in brackets represent the prediction results of one, three, and five steps of one model, respectively). In comparison, the R^2 of VMD-VAE-LSTM are (0.993, 0.980, 0.970). From the analysis above, the R^2 of VMD-VAE-LSTM is higher than that of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM, respectively. MAPE, MAE, and RMSE of VMD-VAE-LSTM were lower than those of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM, respectively.

VMD-VAE-LSTM maintains the lowest prediction error in multi-step prediction. When forecasting the actual power load data from Nanjing, VMD-VAE-LSTM provides better performance for load data forecasting.

4.2. Example 2: Taixing

This section describes an additional evaluation. The performance of VMD-VAE-LSTM in this paper is also evaluated through the electric load data from Taixing. The power load data in Taixing is collected every 24 hours. The data set presents the total electricity consumption of Taixing. The data set in Taixing has 939 data points (from 00:00 in May 13, 2018 to 00:00 on December 5, 2020). The descriptive statistics information of the data set in Taixing is presented in Table 4. The segmentation of the data set reference Section 4.1 is separated into two parts: The training set has 780 data points, the testing set has 159 data points. The actual electric load value and the predicted results based on VMD-VAE-LSTM are presented in Figure 6. The outcomes of the experiment in Example 2 are presented in Table 5. The details of the experiment are provided below.

4.2.1. Comparison among EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM

To verify the improvement of VMD in the forecasting performance of VMD-VAE-LSTM in a different data set, this section adopts Taixing data set. The empirical method is used in VMD to set up the

Table 3: Multi-step forecasting results of VMD-VAE-LSTM with its comparison model in the Nanjing data set.

Multi-step forecasting	Model	R^2	MAPE	MAE	RMSE
One-step-ahead forecasting	RNN	0.757	0.051	110.498	149.446
	LSTM	0.823	0.042	91.882	127.566
	EMD-LSTM	0.881	0.040	85.727	104.427
	EEMD-LSTM	0.908	0.036	78.884	91.984
	CEEMD-LSTM	0.900	0.036	78.131	96.119
	CEEMDAN-LSTM	0.918	0.030	65.996	86.638
	VMD-LSTM	0.930	0.032	69.638	80.092
	VMD-AE-LSTM	0.963	0.026	56.157	58.375
	VMD-VAE-LSTM	**0.993**	**0.010**	**21.194**	**25.747**
Three-step-ahead forecasting	RNN	0.674	0.064	138.664	196.544
	LSTM	0.772	0.049	106.839	144.939
	EMD-LSTM	0.763	0.056	121.453	147.755
	EEMD-LSTM	0.806	0.051	110.514	133.66
	CEEMD-LSTM	0.835	0.044	94.211	123.210
	CEEMDAN-LSTM	0.908	0.034	72.949	92.216
	VMD-LSTM	0.928	0.031	66.045	81.327
	VMD-AE-LSTM	0.957	0.021	47.163	62.670
	VMD-VAE-LSTM	**0.980**	**0.016**	**35.032**	**43.276**
Five-step-ahead forecasting	RNN	0.531	0.076	161.365	235.502
	LSTM	0.728	0.054	116.902	158.026
	EMD-LSTM	0.653	0.068	147.843	178.646
	EEMD-LSTM	0.843	0.044	95.920	120.276
	CEEMD-LSTM	0.649	0.070	150.790	179.587
	CEEMDAN-LSTM	0.892	0.035	75.971	99.795
	VMD-LSTM	0.914	0.031	68.148	88.832
	VMD-AE-LSTM	0.943	0.028	60.994	72.365
	VMD-VAE-LSTM	**0.970**	**0.019**	**41.152**	**52.324**

value of K. Finally, according to several experiments, Taixing data set was decomposed into 9 IMFs and one residual component. As depicted in Figure 5, 412 data points (from 00:00 on May 13, 2018 to 00:00 on June 28, 2019) are selected to show decomposition results. These components have the characteristics of being periodic and regular, which is useful for learning more information from data and promoting forecasting accuracy. In addition, the experimental in this section is designed with several different decomposition algorithms to compare with VMD-LSTM. To show the different performances of

Table 4: The descriptive statistics of the Taixing data set.

data set	Size	Min.	Max.	Median	Mean	Std. Dev.
Total data	939	1210.872	2578.343	1842.248	1842.845	216.967
Training data	780	1210.872	2516.598	1808.684	1805.465	198.727
Testing data	159	1605.078	2578.343	1970.314	2026.220	209.808

The unit of load data in Taixing data set is 10MW·h.

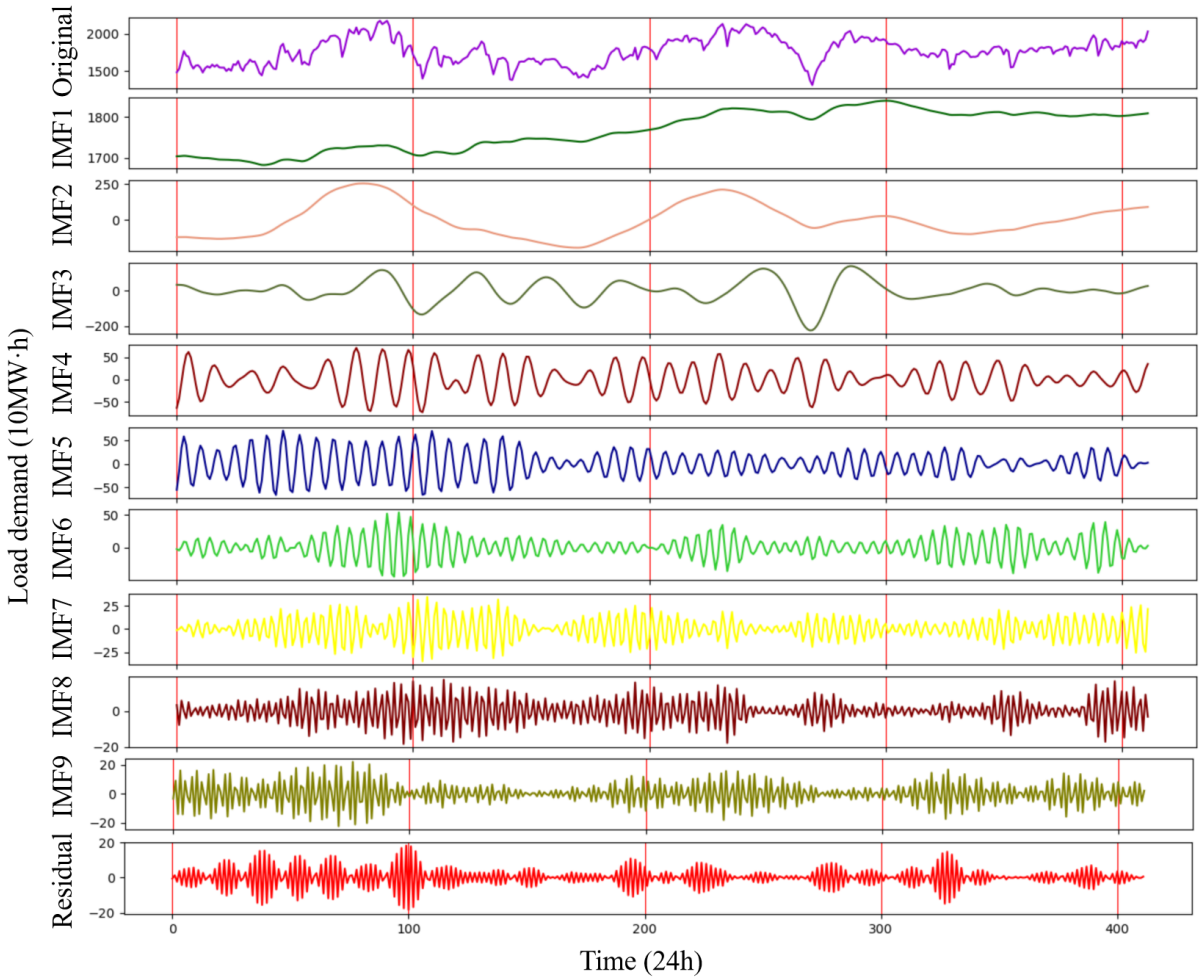


Figure 5: VMD components of the partial Taixing data set (The time of the partial Taixing data set is from 00:00 on May 13, 2018 to 00:00 on June 28, 2019).

EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM, the corresponding power load forecasting curves are presented in Figure 6. The error indexes of prediction performance presented in Table 5 indicate that VMD-VAE-LSTM has the smallest error index and the best fitting degree. They also reveal that VMD-VAE-LSTM has a better forecasting performance than other contrast models.

As indicated in Figure 6, the prediction model using the decomposition algorithm obtains a better fitting degree than LSTM. In Table 5, through the error indexes, the forecasting performance of the model using the decomposition algorithm is better than LSTM and RNN. Therefore, the decomposition algorithm has the advantage of promoting the feature extraction performance of the prediction model.

Furthermore, VMD-LSTM has a better fitting degree than other contrast models. Compared with the model that only uses the decomposition algorithm in Figures 4(a)–4(c), it explains that the power load forecasting curve of VMD-LSTM has the highest fitting degree to the actual power data curve than that of other contrast models. This represents that the forecasting effect of VMD-LSTM is better than other models. Further, VMD can extract more effective feature information. Compared with other comparative decomposition algorithms, VMD is a completely no-recursive variational pattern decomposition model. Therefore, VMD can avoid errors caused during the calculation of recursion and the end of recursion. VMD has a better decomposition performance. Therefore, VMD-LSTM has a great performance in multi-step prediction. Meanwhile, the results presented in Table 5 show that VMD can decompose data better according to frequency and has a wider range of applications. The power load data predicted by VMD-LSTM is more accurate.

4.2.2. Comparison among VMD-LSTM, VMD-AE-LSTM, and VMD-VAE-LSTM

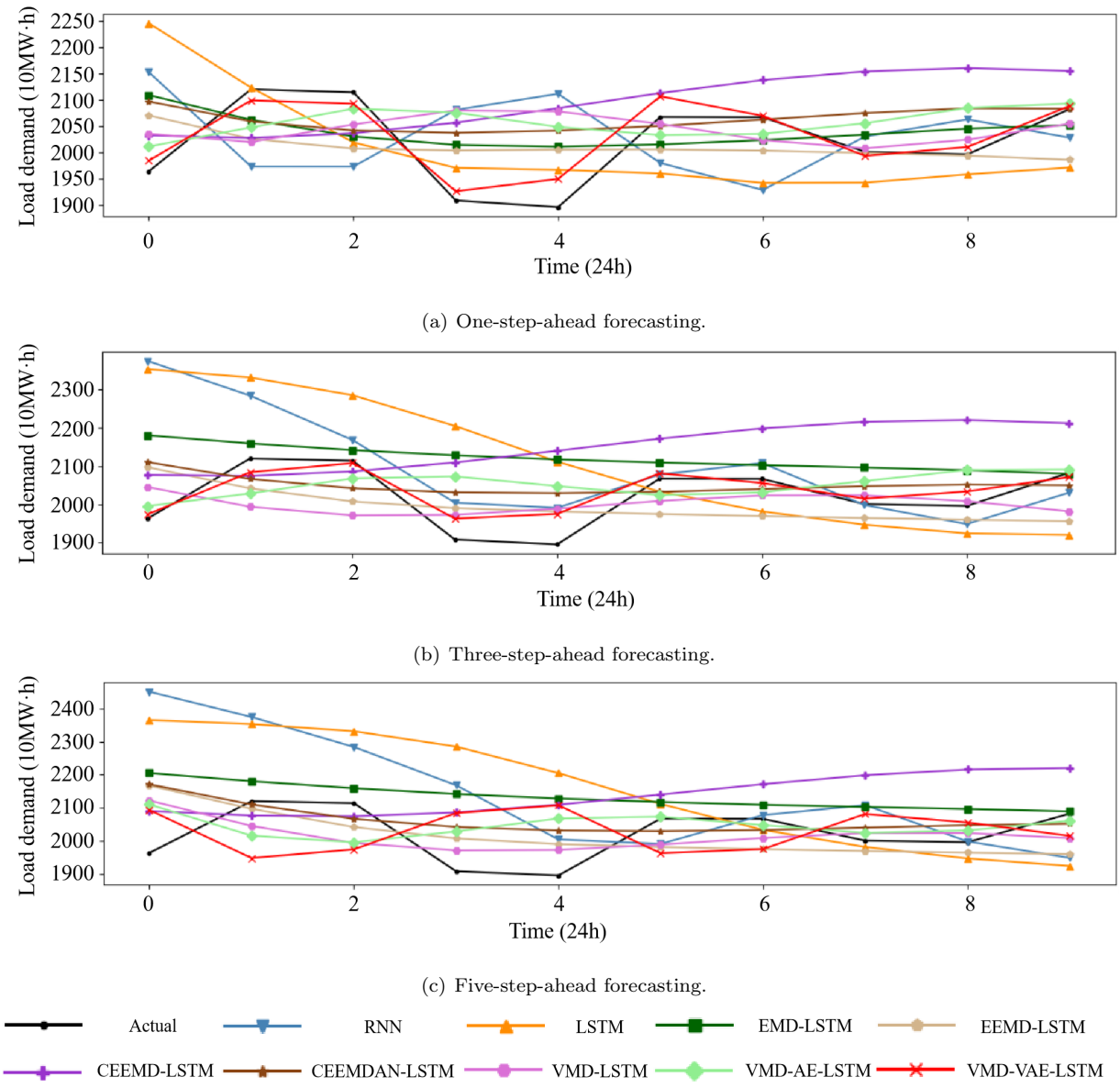
To study the improved performance of VMD-VAE-LSTM by using VAE to reduce the dimensionality of data, this experiment compares VMD-VAE-LSTM with VMD-LSTM and VMD-AE-LSTM. The corresponding prediction results are presented in Figure 6.

VAE can effectively promote the forecasting performance of the model. In Figure 6, compared with the prediction result of VMD-LSTM, the prediction result using VAE is closer to actual data. This implies that dimensionality reduction is beneficial for the performance of the prediction model. Meanwhile, VMD-VAE-LSTM is compared with VMD-AE-LSTM. Compared with AE, VAE can learn smooth latent state representations of the input data. Therefore, VAE can extract more important feature information. As the results are shown in Table 5, VMD-VAE-LSTM maintains the lowest prediction error in multi-step prediction. In actual power load forecasting, this model can achieve a better result of power load forecasting. In brief, VMD-VAE-LSTM has higher forecasting accuracy.

4.2.3. Comparison among EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, VMD-LSTM, and VMD-VAE-LSTM

VMD-VAE-LSTM has better forecasting performance as compared to popular models. We further compare five popular prediction models and the developed model—EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, VMD-LSTM, and VMD-VAE-LSTM.

The comparison reveals that the forecasting result of VMD-VAE-LSTM is closer to the actual power load data than contrast models. In Table 5, the R^2 of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM are (0.876, 0.815, 0.768), (0.901, 0.885, 0.805), (0.859, 0.783, 0.761),



(a) One-step-ahead forecasting.

(b) Three-step-ahead forecasting.

(c) Five-step-ahead forecasting.

Figure 6: One-, three- and five-step-ahead forecasting of various hybrid models in the Taixing data set (The load data of Taixing is from 00:00 on August 29, 2020 to 00:00 on September 8, 2020).

(0.888, 0.862, 0.850), and (0.916, 0.901, 0.898), respectively (The values in parentheses represent the prediction results of one, three, and five steps of one model, respectively). In comparison, the R^2 of VMD-VAE-LSTM are (0.990, 0.977, 0.940). From these results, the R^2 of VMD-VAE-LSTM is higher than that of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM, respectively. Moreover, MAPE, MAE, and RMSE of VMD-VAE-LSTM are lower than those of EMD-LSTM, EEMD-LSTM, CEEMD-LSTM, CEEMDAN-LSTM, and VMD-LSTM, respectively.

In summary, the VMD-VAE-LSTM model designed in this paper forecasts the electric load data from Taixing, which has the characteristics of good robustness and high precision.

Table 5: Multi-step forecasting results of VMD-VAE-LSTM with its comparison model in the Taixing data set.

Multi-step forecasting	Model	R^2	MAPE	MAE	RMSE
One-step-ahead forecasting	RNN	0.752	0.039	77.292	104.166
	LSTM	0.784	0.037	75.086	105.041
	EMD-LSTM	0.876	0.030	59.643	73.701
	EEMD-LSTM	0.901	0.027	54.413	65.890
	CEEMD-LSTM	0.859	0.028	56.818	78.526
	CEEMDAN-LSTM	0.888	0.026	52.130	70.037
	VMD-LSTM	0.916	0.022	43.490	60.437
	VMD-AE-LSTM	0.949	0.018	35.285	47.400
	VMD-VAE-LSTM	**0.990**	**0.008**	**16.960**	**20.668**
Three-step-ahead forecasting	RNN	0.721	0.042	86.800	119.381
	LSTM	0.671	0.046	93.168	129.555
	EMD-LSTM	0.815	0.033	66.853	89.846
	EEMD-LSTM	0.885	0.026	52.192	70.926
	CEEMD-LSTM	0.783	0.037	74.538	97.388
	CEEMDAN-LSTM	0.862	0.029	57.597	77.635
	VMD-LSTM	0.901	0.024	48.529	65.911
	VMD-AE-LSTM	0.932	0.0216	43.404	54.571
	VMD-VAE-LSTM	**0.977**	**0.013**	**25.288**	**31.838**
Five-step-ahead forecasting	RNN	0.501	0.057	116.403	159.631
	LSTM	0.584	0.051	104.321	145.660
	EMD-LSTM	0.768	0.039	78.295	100.723
	EEMD-LSTM	0.805	0.034	68.541	92.303
	CEEMD-LSTM	0.761	0.040	79.507	102.340
	CEEMDAN-LSTM	0.850	0.028	56.840	81.033
	VMD-LSTM	0.898	0.027	53.991	66.912
	VMD-AE-LSTM	0.905	0.027	53.439	64.379
	VMD-VAE-LSTM	**0.940**	**0.021**	**42.125**	**51.226**

5. Conclusion

An accurate and effective power load forecasting model plays a significant role in the safe operation of a power grid. Nevertheless, the complexity of power load data and extraction of effective features brings challenges to prediction. The existing prediction models cannot efficiently solve these challenges. Therefore, the VMD-VAE-LSTM model has been developed in this paper to improve the accuracy of electric

load forecasting. The model combines LSTM with VMD and VAE. VMD is used for the decomposition of power load data, and VAE is used for processing complex data. The feature extraction module can provide more prominent feature information for sub-sequences. In addition, this paper performs multi-step forecasting on two real power load data sets in Nanjing and Taixing and compares the VMD-VAE-LSTM model with the other 7 models. The mean absolute percentage error values for one-step forecasts are 1% and 0.8%, respectively. In conclusion, VMD-VAE-LSTM shows great advantages in processing complex power load forecasting data and improving forecasting accuracy.

Although the VMD-VAE-LSTM model can achieve noteworthy advances, there still exist a few constraints, such as error accumulation when using VAE for feature extraction. In the future, we will apply VMD-VAE-LSTM model to other time sequences prediction problems. In addition, differently from load forecasting based on time-series, considering a few related external factors (such as temperature and holidays) can have the potential to further improve forecasting accuracy. Thus, incorporating these external factors for load forecasting can be further explored as a direction for future research.

Declaration of interests

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.

Credit authorship contribution Statement

Yang Yang: Writing - review & editing, Funding acquisition. Zijin Wang: Software, Visualization, Formal analysis, Writing-original draft. Yuchao Gao: Writing-original draft. Jinran Wu: Supervision, Formal analysis, Writing-original draft, Writing-review & editing. Shangrui Zhao: Writing-review & editing. Zhe Ding: Writing-review & editing.

Acknowledgements

The work is supported by the Australian Research Council project (grant number DP160104292), the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS) (grant number CE140100049) and the Chinese Fundamental Research Funds for the Central Universities (WUT: 213114009). This work is supported in part by the National Natural Science Foundation of China under Grant 61873130 and Grant 61833011, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191377, in part by the 1311 Talent Project of Nanjing University of Posts and Telecommunications.

Appendix A. The parameter setting of the benchmark model

Appendix B. Evaluation metrices

Table A.6: Parameter setting of the benchmark model.

Example	Model	Hyperparametric			
		Hidden neurons	Learning rate (α)	epochs	K (the number of components)
Example 1 (Naijing)	RNN	(256, 128, 64, 1)	0.01	100	-
	LSTM	(256, 128, 64, 1)	0.01	100	-
	EMD-LSTM	(256, 128, 64, 1)	0.01	100	10
	EEMD-LSTM	(256, 128, 64, 1)	0.01	100	15
	CEEMD-LSTM	(256, 128, 64, 1)	0.01	100	9
	CEEMDAN-LSTM	(256, 128, 64, 1)	0.01	100	15
	VMD-LSTM	(256, 128, 64, 1)	0.01	100	10
Example 2 (Taixing)	RNN	(256, 128, 64, 1)	0.01	100	-
	LSTM	(256, 128, 64, 1)	0.01	100	-
	EMD-LSTM	(256, 128, 64, 1)	0.01	100	7
	EEMD-LSTM	(256, 128, 64, 1)	0.01	100	10
	CEEMD-LSTM	(256, 128, 64, 1)	0.01	100	9
	CEEMDAN-LSTM	(256, 128, 64, 1)	0.01	100	10
	VMD-LSTM	(256, 128, 64, 1)	0.01	100	10

Table B.7: Evaluation metrices.

Metrics	Full name	Expression
MAE	Mean absolute error	$\text{MAE} = \frac{1}{S}\sum_{i=1}^{S}\lvert P_i - \hat{P}_i\rvert$
RMSE	Root mean square error	$\text{RMSE} = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(P_i - \hat{P}_i)^2}$
MAPE	Absolute percentage error	$\text{MAPE} = \frac{1}{S}\sum_{i=1}^{S}\lvert\frac{100\times(P_i-\hat{P}_i)}{P_i}\rvert$
R^2	Goodness-of-fit	$R^2 = 1 - \sum_{i=1}^{S}(P_i - \hat{P}_i)^2 / \sum_{i=1}^{S}(P_i - \frac{1}{S}\sum_{i=1}^{S}P_i)$

References

[1] Akshita Gupta, Arun Kumar, and K. Boopathi. Intraday wind power forecasting employing feedback mechanism. *Electric Power Systems Research*, 201:107518, 2021.

[2] Zhe Chen, Yongbao Chen, Tong Xiao, Huilong Wang, and Pengwei Hou. A novel short-term load forecasting framework based on time-series clustering and early classification algorithm. *Energy and Buildings*, 251:111375, 2021.

[3] Dengji Zhou, Shixi Ma, Jiarui Hao, Dong Han, Dawen Huang, Siyun Yan, and Taotao Li. An

electricity load forecasting model for integrated energy system based on bigan and transfer learning. *Energy Reports*, 6:3446–3461, 2020.

[4] Xueheng Qiu, Ye Ren, Ponnuthurai Nagaratnam Suganthan, and Gehan AJ Amaratunga. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing*, 54:246–255, 2017.

[5] Fatemeh Moinian and Mohammad Taghi Ameli. A reliability-based approach for integrated generation and transmission maintenance coordination in restructured power systems. *Electric Power Systems Research*, 206:107737, 2022.

[6] Lavinia Marina Paola Ghilardi, Alessandro Francesco Castelli, Luca Moretti, Mirko Morini, and Emanuele Martelli. Co-optimization of multi-energy system operation, district heating/cooling network and thermal comfort management for buildings. *Applied Energy*, 302:117480, 2021.

[7] Ghulam Hafeez, Imran Khan, Sadaqat Jan, Ibrar Ali Shah, Farrukh Aslam Khan, and Abdelouahid Derhab. A novel hybrid load forecasting framework with intelligent feature engineering and optimization algorithm in smart grid. *Applied Energy*, 299:117178, 2021.

[8] Liye Xiao, Wei Shao, Tulu Liang, and Chen Wang. A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting. *Applied Energy*, 167:135–153, 2016.

[9] Yongquan Dong, Zichen Zhang, and Wei-Chiang Hong. A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. *Energies*, 11(4):1009, 2018.

[10] Jian Luo, Tao Hong, and Shu-Cherng Fang. Robust regression models for load forecasting. *IEEE Transactions on Smart Grid*, 10(5):5397–5404, 2018.

[11] Ziteng Wen, Linbo Xie, Qigao Fan, and Hongwei Feng. Long term electric load forecasting based on ts-type recurrent fuzzy neural network model. *Electric Power Systems Research*, 179:106106, 2020.

[12] Nazih Abu-Shikhah and Fawwaz Elkarmi. Medium-term electric load forecasting using singular value decomposition. *Energy*, 36:4259–4271, 2011.

[13] Azim Heydari, Meysam Majidi Nezhad, Elmira Pirshayan, Davide Astiaso Garcia, Farshid Keynia, and Livio De Santoli. Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm. *Applied Energy*, 277: 115503, 2020.

[14] Wenquan Xu, Hui Peng, Xiaoyong Zeng, Feng Zhou, Xiaoying Tian, and Xiaoyan Peng. A hybrid modelling method for time series forecasting based on a linear regression model and deep learning. *Applied Intelligence*, 49(8):3002–3015, 2019.

[15] Bangru Xiong, Lu Lou, Xinyu Meng, Xin Wang, Hui Ma, and Zhengxia Wang. Short-term wind power forecasting based on attention mechanism and deep learning. *Electric Power Systems Research*, 206:107776, 2022.

[16] Yi Wang, Gabriela Hug, Zijie Liu, and Ning Zhang. Modeling load forecast uncertainty using generative adversarial networks. *Electric Power Systems Research*, 189:106732, 2020.

[17] Michel Bessani, Julio A.D. Massignan, Talysson M.O. Santos, João B.A. London, and Carlos D. Maciel. Multiple households very short-term load forecasting using bayesian networks. *Electric Power Systems Research*, 189:106733, 2020.

[18] Mohamed El-Hendawi and Zhanle Wang. An ensemble method of full wavelet packet transform and neural network for short term electrical load forecasting. *Electric Power Systems Research*, 182: 106265, 2020.

[19] Yang Yang, Hu Zhou, Jinran Wu, Zhe Ding, and You-Gan Wang. Robustified extreme learning machine regression with applications in outlier-blended wind-speed forecasting. *Applied Soft Computing*, 122:108814, 2022.

[20] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01):1–41, 2009.

[21] Junrong Zhang, Huiming Tang, Dwayne D Tannant, Chengyuan Lin, Ding Xia, Xiao Liu, Yongquan Zhang, and Junwei Ma. Combined forecasting model with ceemd-lcss reconstruction and the abc-svr method for landslide displacement prediction. *Journal of Cleaner Production*, 293:126205, 2021.

[22] VSRP Varma Kamadi, Appa Rao Allam, Sita Mahalakshmi Thummala, et al. A computational intelligence technique for the effective diagnosis of diabetic patients using principal component analysis (pca) and modified fuzzy sliq decision tree approach. *Applied Soft Computing*, 49:137–145, 2016.

[23] Shih-Wei Lin and Shih-Chieh Chen. Psolda: A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis. *Applied Soft Computing*, 9(3):1008–1015, 2009.

[24] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *2th International Conference on Learning Representations, ICLR*, 2014.

[25] Marco Inacio, Rafael Izbicki, and Bálint Gyires-Tóth. Distance assessment and analysis of high-dimensional samples using variational autoencoders. *Information Sciences*, 557:407–420, 2021.

[26] Chen Zhang, Zhuo Tang, Youfei Zuo, Kenli Li, and Keqin Li. A robust generative classifier against transfer attacks based on variational auto-encoders. *Information Sciences*, 550:57–70, 2021.

[27] Qingyang Xu, Zhe Wu, Yiqin Yang, and Li Zhang. The difference learning of hidden layer between autoencoder and variational autoencoder. In *2017 29th Chinese Control And Decision Conference (CCDC)*, pages 4801–4804. IEEE, 2017.

[28] Ryno Laubscher and Pieter Rousseau. An integrated approach to predict scalar fields of a simulated turbulent jet diffusion flame using multiple fully connected variational autoencoders and mlp networks. *Applied Soft Computing*, 101:107074, 2021.

[29] M Lydia, S Suresh Kumar, A Immanuel Selvakumar, and G Edwin Prem Kumar. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Conversion and Management*, 112:115–124, 2016.

[30] Siddharth Arora and James W Taylor. Rule-based autoregressive moving average models for forecasting load on special days: A case study for france. *European Journal of Operational Research*, 266(1):259–268, 2018.

[31] Qin Shu, Yu Fan, Fangwei Xu, Chang Wang, and Jinyan He. A harmonic impedance estimation method based on ar model and burg algorithm. *Electric Power Systems Research*, 202:107568, 2022.

[32] M Hadi Amini, Amin Kargarian, and Orkun Karabasoglu. Arima-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electric Power Systems Research*, 140:378–390, 2016.

[33] KE ArunKumar, Dinesh V Kalaga, Ch Mohan Sai Kumar, Govinda Chilkoor, Masahiro Kawaji, and Timothy M Brenza. Forecasting the dynamics of cumulative covid-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-regressive integrated moving average (arima) and seasonal auto-regressive integrated moving average (sarima). *Applied Soft Computing*, 103:107161, 2021.

[34] Omar H Abdalla, Maged A Abu Adma, and Abdelmonem S Ahmed. Two-stage robust generation expansion planning considering long-and short-term uncertainties of high share wind energy. *Electric Power Systems Research*, 189:106618, 2020.

[35] R Le Goff Latimier, E Le Bouedec, and V Monbet. Markov switching autoregressive modeling of wind power forecast errors. *Electric Power Systems Research*, 189:106641, 2020.

[36] S Sp Pappas, L Ekonomou, P Karampelas, DC Karamousantas, SK Katsikas, GE Chatzarakis, and PD Skafidas. Electricity demand load forecasting of the hellenic power system using an arma model. *Electric Power Systems Research*, 80(3):256–264, 2010.

[37] Sebastian Maldonado, Agustin Gonzalez, and Sven Crone. Automatic time series analysis for electric load forecasting via support vector regression. *Applied Soft Computing*, 83:105616, 2019.

[38] Xiaobo Zhang, Jianzhou Wang, and Kequan Zhang. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm. *Electric Power Systems Research*, 146:270–285, 2017.

[39] Tae-Young Kim and Sung-Bae Cho. Predicting residential energy consumption using cnn-lstm neural networks. *Energy*, 182:72–81, 2019.

[40] Xianlun Tang, Hongxu Chen, Wenhao Xiang, Jingming Yang, and Mi Zou. Short-term load forecasting using channel and temporal attention based temporal convolutional network. *Electric Power Systems Research*, 205:107761, 2022.

[41] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, volume 2, pages 985–990. IEEE, 2004.

[42] Yang Yang, Zhenghang Tao, Chen Qian, Yuchao Gao, Hu Zhou, Zhe Ding, and Jinran Wu. A hybrid robust system considering outliers for electric load series forecasting. *Applied Intelligence*, 52:1630–1652, 2022.

[43] Yang Yang, Hu Zhou, Yuchao Gao, Jinran Wu, You-Gan Wang, and Liya Fu. Robust penalized extreme learning machine regression with applications in wind speed forecasting. *Neural Computing and Applications*, 34:391–407, 2021.

[44] Jinran Wu, You-Gan Wang, Yu-Chu Tian, Kevin Burrage, and Taoyun Cao. Support vector regression with asymmetric loss for optimal electric load forecasting. *Energy*, 223:119969, 2021.

[45] Neeraj Neeraj, Jimson Mathew, Mayank Agarwal, and Ranjan Kumar Behera. Long short-term memory-singular spectrum analysis-based model for electric load forecasting. *Electrical Engineering*, 103(2):1067–1082, 2021.

[46] Gong Cheng, Xinzhi Wang, and Yurong He. Remaining useful life and state of health prediction for lithium batteries based on empirical mode decomposition and a long and short memory neural network. *Energy*, 232:121022, 2021.

[47] Rabin K Jana, Indranil Ghosh, and Manas K Sanyal. A granular deep learning approach for predicting energy consumption. *Applied Soft Computing*, 89:106091, 2020.

[48] Vijay Krishna, SP Mishra, Jyotirmayee Naik, and PK Dash. Adaptive vmd based optimized deep learning mixed kernel elm autoencoder for single and multistep wind power forecasting. *Energy*, page 122585, 2021.

[49] Jun Li, Shuqing Zhang, and Zhenning Yang. A wind power forecasting method based on optimized decomposition prediction and error correction. *Electric Power Systems Research*, 208:107886, 2022.

[50] Tingting Zhang, Zhenpeng Tang, Junchuan Wu, Xiaoxu Du, and Kaijie Chen. Short term electricity price forecasting using a new hybrid model based on two-layer decomposition technique and ensemble learning. *Electric Power Systems Research*, 205:107762, 2022.

[51] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3):531–544, 2013.

[52] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

[53] Adam B Levy. Convergence of successive approximation methods with parameter target sets. *Mathematics of Operations Research*, 30(3):765–784, 2005.

[54] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. Recurrent neural network regularization. In *3th International Conference on Learning Representations, ICLR*, 2015.

[55] María E Torres, Marcelo A Colominas, Gastón Schlotthauer, and Patrick Flandrin. A complete ensemble empirical mode decomposition with adaptive noise. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4144–4147. IEEE, 2011.

[56] Jatin Bedi and Durga Toshniwal. Energy load time-series forecast using decomposition and autoencoder integrated memory network. *Applied Soft Computing*, 93:106390, 2020.