

A VMD and LSTM based hybrid model of load forecasting for power grid security

Lingling Lv, Zongyu Wu, Jinhua Zhang, Lei Zhang, Zhiyuan Tan, *Member, IEEE*, and Zhihong Tian, *Senior Member, IEEE* .

Abstract—As the basis for the static security of the power grid, power load forecasting directly affects the safety of grid operation, the rationality of grid planning, and the economy of supply-demand balance. However, various factors lead to drastic changes in short-term power consumption, making the data more complex and thus more difficult to forecast. In response to this problem, a new hybrid model based on Variational mode decomposition (VMD) and Long Short-Term Memory (LSTM) with seasonal factors elimination and error correction is proposed in this paper. Comprehensive case studies on four real-world load datasets from Singapore and the United States are employed to demonstrate the effectiveness and practicality of the proposed hybrid model. The experimental results show that the prediction accuracy of the proposed model is significantly higher than that of the contrast models.

Index Terms—Power grid security, short-term load forecasting, seasonal factors elimination, error correction.

I. INTRODUCTION

THE load of the power grid is closely related to the security of the power grid. Short-term load forecasting (STLF) is a necessary condition to ensure the safety and economic operation of the power grid. It is also the basis for formulating power generation and power supply plans and balancing the supply and demand of the power grid. Only accurate load forecasting can reasonably arrange the operation mode of the power grid and maximize the benefits of the power supply enterprises. At the same time, load forecasting is also an effective proof of the maintenance and operation mode of electrical facilities. Therefore, the practicality and accuracy of STLF models are significant to the safety and stability of smart grid normal operations [1].

There is a complex non-linear relationship between electric load data. Therefore, decomposition algorithms are used by scholars to decompose the load data into multiple subsequences

This work is supported by the National Key Research and Development Program of China (No. 2019YFE0104800), Scientific and Technological Innovation Team of Colleges and Universities in Henan Province (No. 22IRTSTHN011), Excellent Youth Fund of Henan Natural Science Foundation (No. 212300410058), Henan High Level Talent Support Plan (No. ZYQR201810138), and Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2019). (Corresponding author: Lei Zhang, Zhihong Tian.)

Lingling Lv, Zongyu Wu and Jinhua Zhang are with Institute of Electric Power, North China University of Water Resources and Electric Power, Zhengzhou 450011, China (e-mail: lingling_lv@163.com; Zongyu_Wu@yeah.net; zhangjhwind@163.com).

Lei Zhang is with Henan Key Laboratory of Big Data Analysis and Processing, Institute of Data and Knowledge Engineering, School of Computer and Information Engineering, Henan University, Kaifeng 475004, China (e-mail: zhanglei@henu.edu.cn).

Zhiyuan Tan is with the School of Computing, Edinburgh Napier University, EH10 5DT Edinburgh, U.K. (e-mail: z.tan@napier.ac.uk).

Zhihong Tian is with the Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China (e-mail: tianzhihong@gzhu.edu.cn)

to reduce the difficulty of prediction. A hybrid method based on empirical modal decomposition (EMD) and LSTM with extreme learning machine optimization algorithm was proposed for predicting biofuel in [2]. A feature extraction STLF model based on EMD and improved generalized regression neural network (GRNN) with minimizing redundancy and maximizing correlation (mRMR) was presented in [3]. A load identification method based on modified ensemble empirical mode decomposition (MEEMD) for load state identification was proposed in [4]. But there are two technical problems with the EMD-like decomposition algorithm. First, the number of their decompositions is unstable when decomposing complex time series, and the subsequent compilation of the program cannot determine the number of predictive models will lead to interruptions. Second, they will add white noise in the decomposition, which will increase the difficulty of prediction and decrease the prediction accuracy.

The experimental results in [5] show that VMD has better decomposition than CEEMD. A new hybrid short-term prediction model based on VMD-WT (Wavelet transform) and Radial basis function (RBF) for wind speed was proposed in [6]. With VMD decomposing the load data into linear and nonlinear, which contributes to the well prediction accuracy [7]. The VMD model can decompose various continuous time series data, but the problem is how to find the best number of decomposition modes [8]. To solve this problem, many scholars have applied different evaluation methods to consider the number of decomposition modes. Permutation Entropy (PE) was used to determine the number of MEEMD decomposition time series submodels in [9]. The multi-scale alignment entropy (MPE) was used to select the intrinsic mode functions (IMFs) with the highest energy by the authors of [10]. Evaluating the effectiveness of the VMD model for decomposing data depends largely on the purpose of the experiment. Therefore, selecting an appropriate evaluation method is critical for the VMD model.

With the growth of performance of computer, deep learning methods have been used in all kinds of fields as emotion analysis [11], intelligent traffic [12], ensemble classification [13], smart grid and so on. The accuracy of the hybrid methods is improved greatly since which takes advantage of the possibilities of the individual approach [14]. A prediction accuracy of 99.07% is achieved by a novel method based Multidirectional LSTM proposed by the authors of [1]. An effective hybrid model based on VMD and LSTM considering temperature and humidity with Bayesian optimization algorithm for STLF was proposed by the authors of [15]. A building STLF framework based LSTM-like neural networks and attention mechanisms was developed in [16], the prediction results of LSTM-like models with attention were more accurate compared to other LSTM-like models without attention. A composite model based on

TABLE I A summary of the related forecasting works based on deep learning methods

| Work | Deep learning method | Application | Data analysis | Error analysis |
|-----------------|----------------------------|--------------------------------------|---------------|----------------|
| [1] | Multi-LSTM | Forecasting stability of grid | yes | no |
| [15] | VMD+LSTM | Forecasting electric load | yes | no |
| [16] | LSTM/BiLSTM with attention | Forecasting electric load | yes | no |
| [17] | SVM with SecRPSO | Forecasting electric load | yes | no |
| [18] | ILMD+ANN | Forecasting electric load | yes | no |
| [19] | ICEEMD+ARIMA/ELM et al | Forecasting wind speed | yes | no |
| [20] | LSTM with Haar wavelet | Forecasting building energy consumer | yes | no |
| Our work | VMD+LSTM | Forecasting electric load | yes | yes |

SVR considering multi-features especially real-time prices with second-order oscillation and repulsion particle swarm optimization (SecRPSO) for electric load forecasting was proposed in [17]. Combining the improved local mean decomposition (ILMD) method with artificial neural network (ANN) was used to predict the short-term wind speed in [18]. A hybrid wind speed prediction method based on improved CEEMDAN and ANN with a multi-objective optimization algorithm was presented in [19]. A forecasting model based on Haar wavelet and LSTM with the improved sine cosine optimization algorithm for predicting the building energy consumption was presented in [20]. A summary of the above related hybrid methods forecasting work is presented in Table I. Although the above hybrid models achieved good prediction accuracy, they did not take into account data preprocessing for data characteristics and validating the practicality of the models with the latest load data.

In order to overcome the above problems and achieve higher prediction accuracy, a new data pre-processing method which is based on VMD and seasonal factors elimination with error correction is proposed by us. The main contributions of this paper are the following:

- 1) A novel prediction method with detailed mathematical analysis process for achieving high accuracy of short-term electric load prediction is presented in this paper.
- 2) In the process of data pre-processing, the method proposed by us that can reflect the most suitable number of decompositions of VMD, eliminate seasonal factors based holiday and workday features of the original data, examine the data features lost during data pre-processing and recover the lost features.
- 3) Twelve experiments are conducted on the basis of real-world data from the United States and Singapore over the past few months to demonstrate the effectiveness and usefulness of the proposed approach.

The rest of this paper consists of the following: Section II first describes the framework of our proposed method. Section III introduces the whole process of the proposed prediction model, including VMD, PE, seasonal factors elimination, error correction and LSTM. The evaluation metrics are shown in Section IV. In Section V, the case studies on real-world load data from different data sets are used to test the method proposed in this paper. Finally, Section VI concludes the study and discusses future work.

II. MAIN STEPS AND FRAMEWORK

The main methods applied in this paper includes VMD, PE, seasonal factors elimination, error correction and LSTM

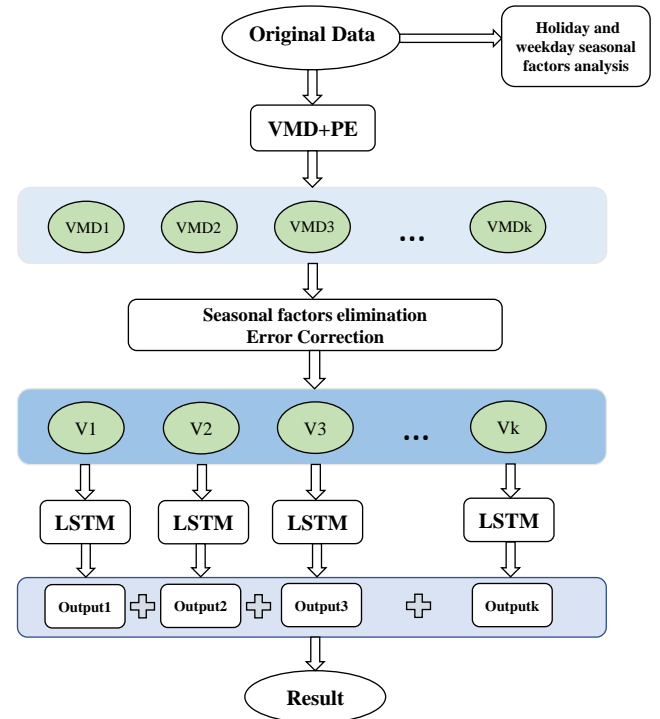


Fig. 1 The whole process of the proposed forecasting model.

as shown in Fig.1. It can be summarized in the following four steps:

- 1) VMD is used to decompose the load data into subsequences, and PE is used to determine the number of subsequences decomposed by the VMD.
- 2) Considering the characteristics of the raw load data, a seasonal factor elimination treatment for weekday and holiday is proposed. Error correction is used to restore data features lost during VMD decomposition and seasonal factors elimination.
- 3) Combine the subsequences with LSTM to forecast the electric load. Different subsequences of the VMD decomposition have different characteristics, so each subsequence is predicted based on a LSTM model with different hyperparameters.
- 4) The final prediction results are obtained by summing the predictions of each LSTM model, and the real data are used to compare with it.

III. MATERIALS AND METHODS

A. VMD(Variational mode decomposition)

VMD is an adaptive, completely non-recursive approach to modal change signal processing. The original load data has complex nonlinear relationships, and using it as input data

for prediction is difficult. VMD can decompose the original load data into multiple simple sequences, which will reduce the difficulty of prediction and improve the prediction accuracy. Different subsequences have different sparse properties, so the determination of the central pulsation and bandwidth of the subsequence are very important in the VMD decomposition. The unilateral frequency spectrum of the subsequence is obtained by Hilbert transform, then the center frequency of each subsequence is estimated by hybrid exponential tuned to shift it to the baseband, and finally the bandwidth of the subsequence is obtained by H^1 Gaussian smoothing [21]. Then the resulting constrained variational problem can be generalized as Eq.(1) :

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

where u_k means the k -th mode, ω_k is the k -th mode center frequency, K means the total number of the modes, and $\delta(t)$ represents the Dirac distribution.

In order to make the above constrained variational problem unconstrained, the combination of quadratic penalty term and Lagrangian multiplier λ is introduced as:

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

The original minimization problem can be solved by the iteration count optimization method which is called alternate direction method of multipliers (ADMM). The u_k and ω_k can be updated by ADMM as Eq.(3), Eq.(4):

1) minimizing u_k :

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

2) minimizing ω_k :

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{\mu}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{\mu}_k(\omega)|^2 d\omega} \quad (4)$$

where n is the number of original data, $\hat{f}(\omega)$, $\hat{\mu}(\omega)$ and $\hat{\lambda}(\omega)$ are the Fourier transform of $f(t)$, $\mu(t)$ and $\lambda(t)$, respectively.

B. Permutation Entropy

The PE (Permutation Entropy) is a measure of the complexity of time series. The description of time series through symbolic space can clearly reflect the correlation between time series data. The embedding size and delay time are key factors in the calculation of the PE. A time series $\{x(i), i = 1, 2, \dots\}$ can be shown as:

$$X_i = [x(i), x(i+L), \dots, x(i+(m-1)L)] \quad (5)$$

where m means embedding dimension, and L means the delay time.

TABLE II PE values for different number of subsequences for VMD decomposition of Singapore datasets.

| Month | Number | | | | | | | |
|-----------|--------------|--------------|--------------|-------|-------|-------|-------|--------------|
| | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| September | 9.164 | 9.340 | 9.391 | 9.448 | 9.612 | 9.557 | 9.521 | 9.122 |
| October | 8.859 | 9.023 | 8.690 | 8.781 | 9.250 | 9.154 | 9.303 | 9.107 |
| November | 8.968 | 9.130 | 8.925 | 9.188 | 9.414 | 9.375 | 9.248 | 9.320 |
| December | 8.901 | 8.329 | 8.865 | 9.184 | 9.028 | 8.916 | 9.109 | 9.327 |
| January | 8.594 | 8.846 | 8.969 | 8.874 | 8.881 | 8.883 | 9.568 | 9.648 |
| February | 8.923 | 8.936 | 8.610 | 8.887 | 8.849 | 8.948 | 8.946 | 9.112 |

In order to find the degree of correlation between the data, each embedding dimension of data which means the m number of X_i is arranged incrementally. The process is shown as follows:

$$X_i = [x(i+(j_1-1)L) \leq x(i+(j_2-1)L) \leq \dots \leq x(i+(j_m-1)L)] \quad (6)$$

where j_i indicates the position of the data before being arranged. Then X_i is mapping onto $[j_1, j_2, \dots, j_m]$, which is one of the $m!$ permutations of m distinct symbols $[1, 2, 3, \dots, m]$.

Introduce a probability distribution P_1, P_2, \dots, P_K to represent the frequency of occurrence of each row X_i in the order of arrangement, where $K \leq m!$. The PE of the time series $\{x(i), i = 1, 2, \dots\}$ can be obtained by calculating the Shannon entropy of K distinct symbols as:

$$H(m) = - \sum_{i=1}^K P_i \ln P_i \quad (7)$$

The experimental results show that the more regular the time series, the smaller the value of the PE; the more complex the time series, the larger the value of the PE.

In this paper, the subseries with different numbers of VMD decompositions are first summed to obtain their time series, and then the ranking entropy of the time series is calculated. In order to avoid the over-decomposition and under-decomposition of VMD decomposition, the minimum number of VMD decomposition is 5 and the maximum is 12.

Table II lists the number of VMD decomposition subsequences from 5 to 12 for the PE values on the Singapore datasets. It can be clearly seen that October, November and February have the lowest PE with a VMD decomposition subsequence number of 7. However, the September PE is minimum when the number of VMD decomposition subsequences is 12 and the January PE is minimum when the number of VMD decomposition subsequences is 5. The over-decomposition occurs when the number of VMD decomposition subsequences is 12, and the under-decomposition occurs when the number of VMD decomposition subsequences is 5. In order to ensure the best VMD decomposition for each month, the number of VMD decomposition subsequences is selected to 7, and the decomposition results of original data are expressed as $VMD1, VMD2, VMD3, VMD4, VMD5, VMD6, VMD7$, where $VMD1$ has the main characteristics of the original data.

C. Error analysis and processing

The data set used in the paper are from Singapore grid data and the U.S. grid data. Seasonal factors have a significant influence on STLF. Therefore, the matter is that how to analysis the original data and eliminate seasonal components.

Fig.2 shows the real data of Singapore in September 2020. These data are sampled every 30 minutes for a total of 48

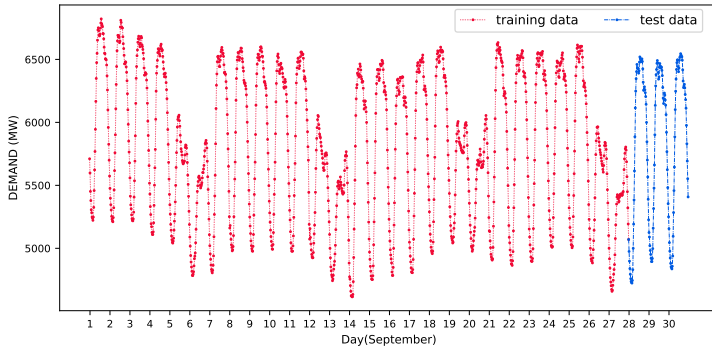


Fig. 2 Electricity demand in Singapore for September 2020.

samples per day. It can be clear seen from the Fig.2 that the maximum load value of September 5 is smaller than September 4. The daily data changes of September 5 and September 6 are similar, but the maximum load value of September 6 is also much smaller than the maximum load value of September 7. The mean values of September 6 and September 7 are also significantly smaller than the mean values of September 5 and September 7. In addition, the variation of the load values from September 1 to September 5 is similar, and the other daily data changes are similar according to a period of 7 days. Therefore, it is perfectly in line with the high load electricity consumption of companies and factories during weekday and low load electricity consumption during holiday. To verify that the seasonal factors of the above data set is a general phenomenon, all data sets from September 2020 to February 2021 are tabulated in this paper as shown in Table III.

TABLE III Weekday and holiday data characteristics for Singapore from September 2020 to February 2021.

| | | Sep. | Oct. | Nov. | Dec. | Jan. | Feb. |
|---------|------|------|------|------|------|------|------|
| Weekday | Min | 4615 | 4844 | 4817 | 4905 | 4617 | 4820 |
| | Max | 6822 | 6839 | 6882 | 6796 | 6680 | 6821 |
| | Mean | 5946 | 6032 | 6031 | 6029 | 5906 | 6041 |
| Holiday | Min | 4658 | 4847 | 4834 | 4865 | 4645 | 4919 |
| | Max | 6056 | 6231 | 6152 | 6169 | 6066 | 6280 |
| | Mean | 5459 | 5583 | 5557 | 5586 | 5471 | 5625 |

Table III clearly shows the minimum, average and maximum values of holiday and weekday for the data set from September 2020 to February 2021 for Singapore. In the September 2020 dataset, the maximum value of 6822 for weekdays is 766 larger than the maximum value of 6056 for holiday, and the average value of 5946 for weekdays is 487 larger than the average value of 5459 for holidays, but the minimum value of 4615 for weekdays is 43 smaller than the minimum value of 4658 for holidays. The difference between the weekday data characteristics and the holiday data characteristics for September 2020 is the largest in all datasets. Therefore, we can conclude that the deviation between the minimum load values for weekday and holiday is small, the deviation between the maximum and average values is large.

In this paper, the data sets are separated into weekday and holiday to reduce the effect of seasonal factors. The data sets are divided into weekday data y and holiday data Y . For weekday data y , the load values at the same moments of each

day are found as the mean value as shown in Eq.(8):

$$\bar{y}_m = \frac{\sum_{i=0}^N y[(n-1) + T^*i]}{N} \quad \forall m \in [0, 1, 2, \dots, M-1] \quad (8)$$

where N means the number of weekday, T is the number of load values for one day, M indicates the number of load data recorded per day. The mean of T moments \bar{y}_m and \bar{y} which is the average value of y are regularized as shown in Eq.(9):

$$\bar{W}_m = \frac{\bar{y}_m}{\bar{y}} \quad \forall m \in [0, 1, 2, \dots, M-1] \quad (9)$$

Then the seasonality index at the m th moment can be obtained:

$$I_m = \frac{\bar{W}_m}{\left(\sum_{n=1}^T \bar{W}_m\right)/T} \quad \forall m \in [0, 1, 2, \dots, M-1] \quad (10)$$

The data set without the effect of seasonal factors y' can be obtained as shown in Eq.(11):

$$y'[n + m * T] = \frac{y[n + m * T]}{I_n} \quad \forall n \in [0, 1, 2, \dots, N] \quad (11)$$

For holiday data Y , the same method is used for processing. After the process of seasonal factors processing, the weekday data and the holiday data are combined to obtain the new data set called *Data*. In the experiment, the seasonality factor treatment process has been done before the VMD decomposition, but the results were not satisfactory. Therefore, the seasonal factor elimination is placed after the VMD decomposition.

D. Error analysis and correction processing

Fast Fourier transform is used in the process of decomposition of the data by VMD, and the noise of the original data is eliminated to some extent. In addition, after the process of seasonal factors mentioned above, the obtained data shows a large error compared to the original data. Fig.3 shows the percentage error of data processed by VMD decomposition and seasonal factors with the original data in Singapore from September 2020 to February 2021. The specific formula for calculating the error is shown in Eq.(12):

$$Error[i] = \frac{(VMD1[i] + VMD2[i] + \dots + VMDk[i]) - data[i]}{data[i]} \quad (12)$$

where $data[i]$ means the i -th point of the original data $data$. From Fig.3 we can find that the largest positive error is 1.9% in December 2020 and the smallest is 1.4% in January 2021. The largest negative error is 2.8% in September 2020 and the smallest is 1.4% in February 2021. Even though the minimum absolute error in February 2021 is only 0.35%, the mean value of 5907 times the absolute error in February 2021 is 20.67, which has a huge bias effect on the final forecast result. When forecasting based on a dataset with large errors, the final prediction accuracy of the model can be poor. Therefore, an error correction process is needed to improve the prediction accuracy.

The error of the data varies from moment to moment, so it is important to calculate the error of each moment with the corresponding moment of the original data for error correction, and the calculation process is shown as:

$$Error^*[i] = \frac{(V1[i] + V2[i] + \dots + Vk[i]) - data[i]}{V1[i] + V2[i] + \dots + Vk[i]} \quad (13)$$

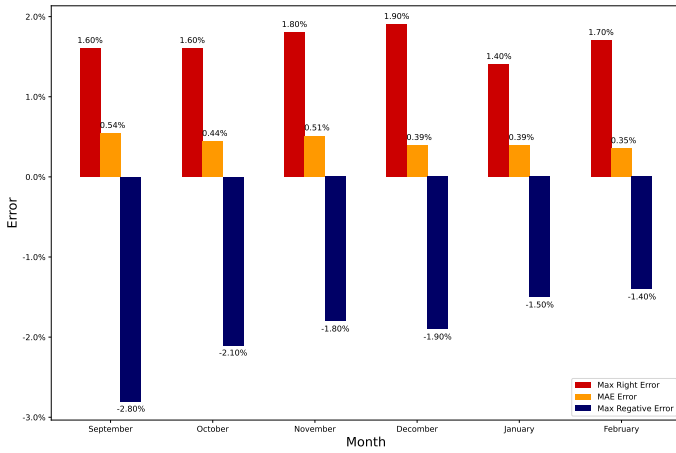


Fig. 3 The percentage error of data processed by VMD decomposition and seasonal factors with the original data in Singapore from September 2020 to February 2021.

where $Error^*[i]$ denotes the error between the original data and the data processed by VMD decomposition and seasonal factors elimination. The data are corrected according to the error at each moment in time, as shown in the following Eq.(14):

$$V_n[i] = \frac{VMD_n[i]}{1 - Error^*[i]} \quad \forall n \in [1, 2, 3, \dots, k] \quad (14)$$

where i denotes the i -th data and $Error^*(\cdot)$ is the deviation of the processed data from the original data divided by the processed data. When $Error^*[i]$ is a negative (positive) number, the processed data can be increased (decreased) after subtraction operation to reduce the deviation from the original data.

E. LSTM(Long Short-Term Memory)

RNN has achieved good results in training short time series in the past, but its hidden layer has only one state, and when dealing with long time series there will be gradient explosion or gradient disappearance problem, which cannot meet the realistic needs. LSTM introduces cell state on the basis of RNN to achieve long-term memory to solve the problems of RNN. LSTM proposes the concept of multiplicative gate cell, through the input gate, forgetting gate and output gate in each storage cell, with the current cell state and the cooperation of sigmoid activation function and tanh activation function, the long time sequence can be predicted and achieved good results. The sigmoid activation function eliminates data less than 0 to achieve the filtering function, while the hyperbolic tangent function adjusts the input information to a number between -1 and 1 to facilitate the calculation. The specific calculation procedure for each cell of LSTM is shown as follows: (1)

- 1) Forget gate: Determine how much useful information from the previous cell state c_{t-1} is retained to the current cell state c_t with the help of the sigmoid activation function.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (15)$$

where $[h_{t-1}, x_t]$ denote combining two vectors into a single vector, W_f denote the weight matrices of forget gate, b_f denote the bias terms of forget gate, and $\sigma(\cdot)$ is the sigmoid function.

- 2) Input gate: The output information h_{t-1} of the previous cell and the input information x_t of the current cell are obtained by the sigmoid activation function and tanh activation function respectively to obtain two different vectors, and then the two different vectors are multiplied to achieve the information input.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (16)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (17)$$

where W_i and W_c denote the weight matrices of input gate and cell state, respectively, b_i and b_c denote the bias terms of input gate and cell state, respectively, and $\tanh(\cdot)$ is the hyperbolic tangent function.

- 3) Cell state: Long-term memory is achieved by retaining previous information and introducing current cellular information at the same time. The state information of the previous cell and the forget gate are dot-producted, and the result of the dot-product operation of the input gate i_t and the current cell state information is added to achieve the state update.

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (18)$$

- 4) Output gate: Realize the screening of all previous cells with valid information. The output information h_{t-1} of the previous cell and the current cell input information x_t are processed by the sigmoid activation function respectively, and then the dot product operation is performed with the output of the cell state c_t processed by the hyperbolic tangent function.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (19)$$

$$h_t = o_t * \tanh(c_t) \quad (20)$$

where W_o denote the weight matrices of output gate, respectively, b_o denote the bias terms of output gate.

IV. PERFORMANCE INDICATORS

In this paper, the mean absolute error (MAE), root mean square error (RMSE), adjusted R-Square (R^2_{adj}) and mean absolute percentage error (MAPE) are selected to evaluate the prediction results of the model. MAE can reflect the absolute value error of each data over the real data, which is the most widely used performance index currently. RMSE provides a visual representation of the predictive effect of the model. $R^2_{adjusted}$ is used to indicate the predictive fit of the model. MAPE can clearly indicate the predictive effect of the model as a percentage of the data and facilitate the analysis of its economic value. These metrics are defined as follows:

- 1) MAE is defined as follows:

$$MAE = \frac{1}{m} \sum_{i=0}^m |Pata[i] - data[i]| \quad (21)$$

- 2) RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=0}^m |Pata[i] - data[i]|^2} \quad (22)$$

3) R_{adj}^2 is defined as follows:

$$R_{adj}^2 = 1 - \frac{\sum_{i=0}^m (Pata[i] - data[i])^2}{\sum_{i=0}^m (\overline{Pata} - data[i])^2} \quad (23)$$

4) MAPE is defined as follows:

$$MAPE = \frac{100}{m} \sum_{i=0}^m \left| \frac{Pata[i] - data[i]}{data[i]} \right| \quad (24)$$

In the above formulas m denotes the data length, $Pata$ denotes the final forecasting results, $data$ denotes the original data, $Pata[i]$ and $data[i]$ denote the values of the data set at a given moment, and \overline{Pata} denotes the average value of the predicted data set.

V. EXPERIMENT AND ANALYSIS

In this section, in order to demonstrate the practicality of the method proposed in this paper, data from two different regions, Singapore and the United States, are selected for forecasting in the most recent months. Four models, ARIMA, CEEMD+LSTM, VMD+LSTM and GRU+RNN [22] are selected for comparison with the model proposed in this paper. Since the IMFs decomposed by VMD and CEEMD decomposition have different characteristics, the corresponding LSTM hyperparameters are also different. After extensive experiments, the input data length of the input LSTM is between 22 and 70, and the maximum computational complexity of the proposed model in this paper is $O(70)$. The validity and practicality of the proposed model in this paper are well verified. All training and analysis for this paper are done on a computer (CPU: Intel Core i7-10700, memory: 16 GB, GPU: Nvidia GeForce RTX 3070). All predictions in this paper are implemented in the Python 3.8 environment, using the TensorFlow 2.4.0 GPU version as the framework.

A. Singapore data specific experiments

In order to verify the practicality of the model proposed in this paper, data for the past six months from September 2020 to February 2021 in Singapore are chosen for this experiment to verify the model's STLF effectiveness. These data sets are sampled every 30 minutes for a total of 48 samples per day. The amount of data for each month is small, so a total of 144 data for the last 3 days of each month is selected as the test set for comparison.

Table IV lists the values of the four error indicators including RMSE, MAE MAPE, and R_{adj}^2 for the proposed and contrast methods. It is clear that the CEEMD+LSTM outperforms ARIMA model, VMD+LSTM model and GRU+RNN model, but it is worse than the model proposed in this paper. Compared with CEEMD+LSTM, the RMSE, MAE, and MAPE of the proposed model in December are decreased by 22.1%, 25.3%, and 26.2%, respectively, which is the least improvement in dll data sets. Fig.4 shows the MAPE and RMSE for the proposed and contrast methods on the Singapore datasets. As we can see in Fig.4 that the MAPE of the model proposed in this paper varies steadily at around 0.15%, which is lower than other forecasting models. Fig.5 shows the MAE and R_{adj}^2 for the proposed and contrast methods on the Singapore datasets. It can be seen from Fig.5 that the purple line is above

TABLE IV Performance evaluation of different models for the Singapore datasets from September 2020 to February 2021.

| Month | Model | Metrics | | | |
|-----------|------------|--------------|-------------|--------------|---------------|
| | | RMSE | MAE | MAPE(%) | R_{adj}^2 |
| September | Proposed | 11.2 | 8.47 | 0.144 | 0.9996 |
| | CEEMD+LSTM | 17.14 | 13.68 | 0.235 | 0.9992 |
| | ARIMA | 29.97 | 24.44 | 0.417 | 0.9976 |
| | VMD+LSTM | 21.84 | 17.26 | 0.292 | 0.9987 |
| | GRU+RNN | 25.8 | 21.14 | 0.364 | 0.9982 |
| October | Proposed | 10.1 | 7.67 | 0.13 | 0.9996 |
| | CEEMD+LSTM | 22.3 | 16.97 | 0.288 | 0.9984 |
| | ARIMA | 31.48 | 23.49 | 0.401 | 0.9968 |
| | VMD+LSTM | 18.52 | 14.36 | 0.241 | 0.9988 |
| | GRU+RNN | 31.66 | 23.32 | 0.391 | 0.9967 |
| November | Proposed | 12.92 | 8.98 | 0.158 | 0.9994 |
| | CEEMD+LSTM | 21.08 | 15.87 | 0.279 | 0.9984 |
| | ARIMA | 31.82 | 23.2 | 0.408 | 0.9964 |
| | VMD+LSTM | 28.97 | 21.13 | 0.367 | 0.9969 |
| | GRU+RNN | 29.6 | 21.35 | 0.377 | 0.9968 |
| December | Proposed | 12.09 | 9.39 | 0.158 | 0.9995 |
| | CEEMD+LSTM | 15.52 | 12.57 | 0.214 | 0.9992 |
| | ARIMA | 33.91 | 26.68 | 0.448 | 0.9964 |
| | VMD+LSTM | 19.47 | 15.22 | 0.255 | 0.9988 |
| | GRU+RNN | 31.54 | 24.66 | 0.417 | 0.9971 |
| January | Proposed | 12.96 | 9.19 | 0.165 | 0.9993 |
| | CEEMD+LSTM | 18.19 | 14.08 | 0.245 | 0.9986 |
| | ARIMA | 32.49 | 24.41 | 0.429 | 0.9957 |
| | VMD+LSTM | 21.95 | 16.8 | 0.295 | 0.9980 |
| | GRU+RNN | 31.59 | 24.4 | 0.423 | 0.9965 |
| February | Proposed | 12.09 | 8.97 | 0.154 | 0.9993 |
| | CEEMD+LSTM | 17.46 | 12.16 | 0.207 | 0.9986 |
| | ARIMA | 29.41 | 22.41 | 0.377 | 0.9961 |
| | VMD+LSTM | 21.37 | 16.24 | 0.275 | 0.9979 |
| | GRU+RNN | 32.04 | 24.40 | 0.412 | 0.9953 |

the pink, blue, orange and green lines, and the purple line indicates the R_{adj}^2 value of the model proposed in this paper, which means that the model proposed in this paper has good accuracy and reliable stability for different data sets. Compared to VMD+LSTM model the MAE of the VMD+LSTM model with seasonal factors elimination and error correction proposed in February 2021 is decreased by 44.8% , which is the smallest improvement in all data sets.

B. The US data specific experiments

To verify that our proposed model remains valid in unstable load datasets, we conducted case studies of real load data from the dataminer2 website of PJM in the US. These data sets include six sets of load values for Mid_Atlantic, PJM_RTO, and Western regions of the US from November 22 in 2020 to January 17 in 2021 and February 1 in 2021 to March 28, which are denoted as 2020 MID, 2020 PJM , 2020 Western, 2021 MID, 2021 PJM, and 2021 Western. These data sets are sampled every hour 24 points a day for a total of 1344 points. These data sets have similar seasonal characteristics to the Singapore data sets in that their weekday load values are greater than their holiday load values as shown in Fig.6. In addition, the data characteristics of the three regions are different. The US data sets are 8 consecutive weeks of data values, and the characteristics of seasonal factors on this basis

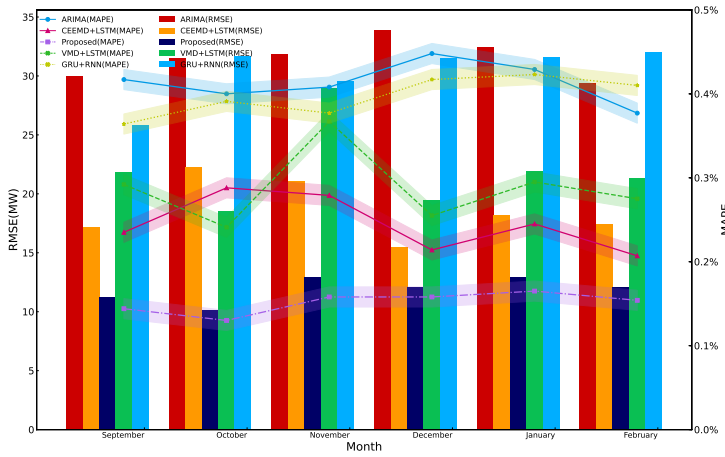


Fig. 4 The PDF graph of MAPE and RMSE of the proposed model and the contrast models for Singapore datasets from September 2020 to February 2021.

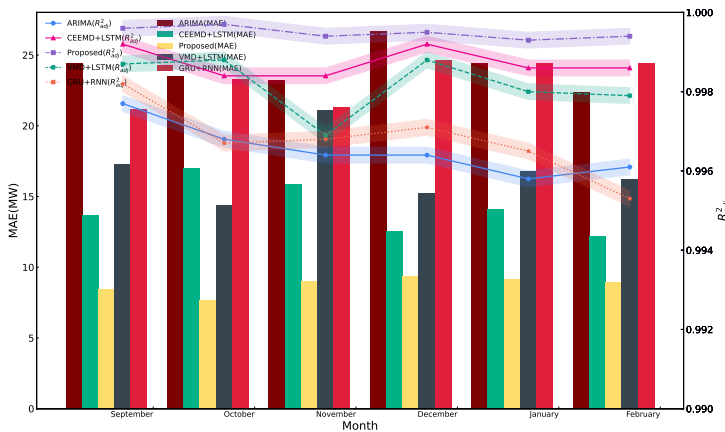


Fig. 5 The PDF graph of MAE and R^2_{adj} of the proposed model and the contrast models for Singapore datasets from September 2020 to February 2021.

are difficult to analyze, which can be clearly found when the data set is divided into 4 weeks. Therefore, the data sets in this subsection are also predicted according to the processing of Singapore data. When decomposing the original data sets by VMD, the PE value of 8 sub-models is less than 7. So the number of sub-models of the VMD for this experiment is 8. Finally, the test set predicted for this experiment is the data from the last 6 days and the length of the test set is 144.

Table V lists the performance evaluations of different methods for the US datasets, where the bolded data are the results of the model proposed in this paper. The RMSE, MAE, MAPE and R^2_{adj} of the proposed method are the smallest among all methods in all datasets, where the MAPE of the proposed model achieved 0.3% in the 2020 PJM dataset. The reason for the larger error in 2021 than in the 2020 dataset is the anomaly in electricity load caused by the impact of a new coronavirus in the US. In Fig. 7, the first three subplots reflect MAE, RMSE and MAPE, and the smallest area is occupied by the error of the proposed model (red line), while the fourth subplot reflects R^2_{adj} , and the largest area is occupied by the proposed model. The experimental results show that the model proposed in this paper has excellent practicality and efficiency.

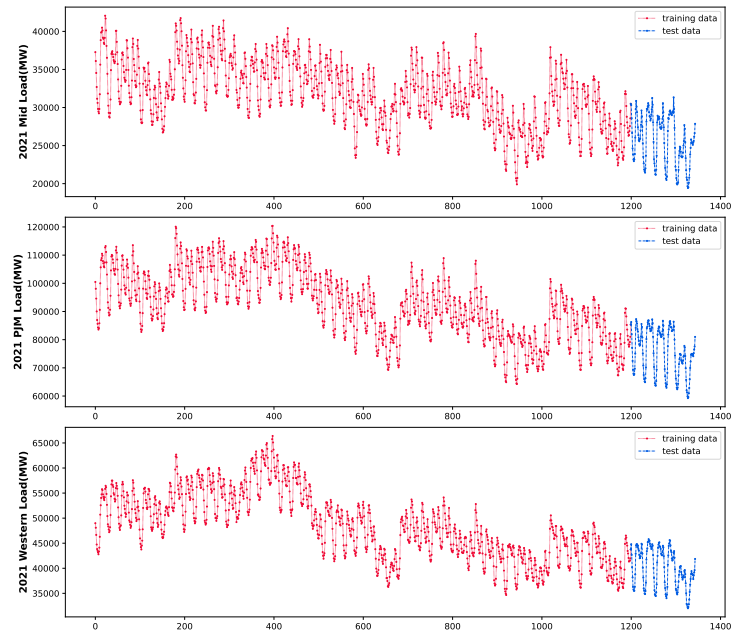


Fig. 6 Electricity demand in the US Mid, PJM and Western regions from February 1 to March 28, 2021.

TABLE V Performance evaluations of different models for the US datasets.

| Distance | Model | Metrics | | | |
|--------------|------------|------------|------------|-------------|---------------|
| | | RMSE | MAE | MAPE(%) | R^2_{adj} |
| 2020 Mid | Proposed | 173 | 121 | 0.4 | 0.9965 |
| | CEEMD+LSTM | 430 | 335 | 1.1 | 0.9799 |
| | ARIMA | 1015 | 725 | 2.3 | 0.8848 |
| | VMD+LSTM | 388 | 309 | 1 | 0.9806 |
| | GRU+RNN | 559 | 426 | 1.3 | 0.9657 |
| 2021 Mid | Proposed | 226 | 187 | 0.7 | 0.9950 |
| | CEEMD+LSTM | 537 | 408 | 1.6 | 0.9760 |
| | ARIMA | 939 | 690 | 2.7 | 0.9141 |
| | VMD+LSTM | 392 | 313 | 1.2 | 0.9840 |
| | GRU+RNN | 747 | 543 | 2.0 | 0.9448 |
| 2020 PJM | Proposed | 379 | 276 | 0.3 | 0.9969 |
| | CEEMD+LSTM | 979 | 792 | 1 | 0.9803 |
| | ARIMA | 2186 | 1498 | 1.6 | 0.9073 |
| | VMD+LSTM | 814 | 648 | 0.7 | 0.9851 |
| | GRU+RNN | 1017 | 733 | 0.8 | 0.9792 |
| 2021 PJM | Proposed | 408 | 336 | 0.45 | 0.9973 |
| | CEEMD+LSTM | 878 | 727 | 0.9 | 0.9886 |
| | ARIMA | 1927 | 1449 | 1.9 | 0.9409 |
| | VMD+LSTM | 846 | 635 | 0.8 | 0.9881 |
| | GRU+RNN | 1398 | 1060 | 1.40 | 0.9683 |
| 2020 Western | Proposed | 242 | 191 | 0.4 | 0.9953 |
| | CEEMD+LSTM | 666 | 520 | 1 | 0.9631 |
| | ARIMA | 1046 | 737 | 1.5 | 0.9157 |
| | VMD+LSTM | 425 | 332 | 0.7 | 0.9845 |
| | GRU+RNN | 544 | 386 | 0.8 | 0.9755 |
| 2021 Western | Proposed | 263 | 186 | 0.5 | 0.9953 |
| | CEEMD+LSTM | 448 | 331 | 0.8 | 0.9866 |
| | ARIMA | 943 | 713 | 1.8 | 0.9388 |
| | VMD+LSTM | 362 | 289 | 0.7 | 0.9905 |
| | GRU+RNN | 697 | 557 | 1.4 | 0.9655 |

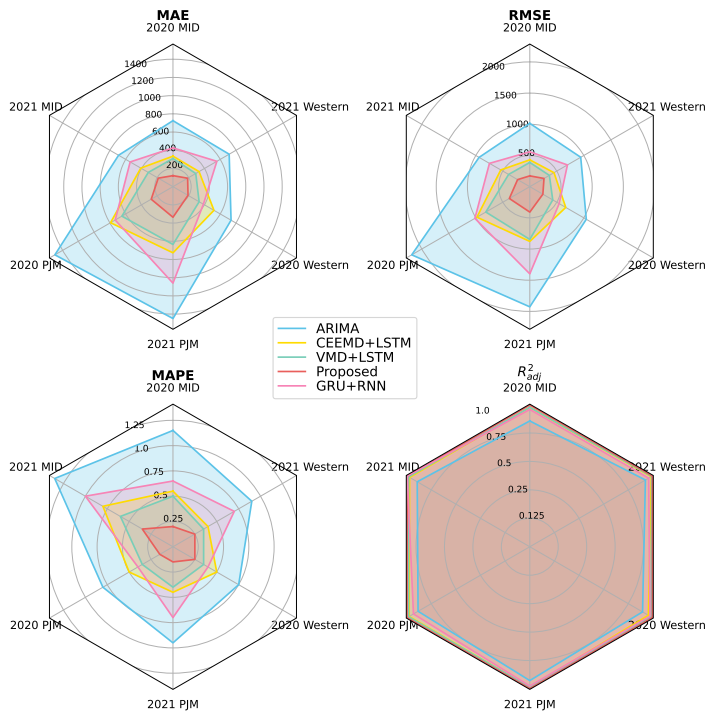


Fig. 7 Polar graph of performance metrics for different forecasting methods based on the US datasets

VI. CONCLUSION

In order to ensure the safety and economy of power grid operation, a new short-term load forecasting model based on VMD, seasonal factors elimination, error correction and LSTM is proposed in this paper. The VMD model is used to reduce the prediction difficulty by decomposing the original complex data into simple sequences and the PE is used to determine the number of decompositions of the VMD. Considering that there is a significant seasonal factors in the electric load data, a seasonal factor elimination process based on the original data characteristics is used to process the electric load data. But some data features are lost after the data has been pre-processed. Therefore, error correction is used to recover lost data features in this paper. In order to fully verify the performance of the model, the data sets selected in this paper are the most recent months of different data characteristics in different regions. The results indicate that the proposed model can obtain the prediction results of the power grid load that meet the real demand. Industry 5.0 is the next industrial evolution, and technologies like digital twins, edge computing and blockchain play a key role in it [23]. In future studies, STLF methods based on deep learning can be combined with edge computing and blockchain to implement smart grid 2.0.

REFERENCES

- [1] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional lstm model for predicting the stability of a smart grid," *IEEE Access*, vol. 8, pp. 85 454–85 463, 2020.
- [2] L. Yu, S. Liang, R. Chen, and K. K. Lai, "Predicting monthly biofuel production using a hybrid ensemble forecasting methodology," *International Journal of Forecasting*, 2019.
- [3] Y. Liang, D. Niu, and W.-C. Hong, "Short term load forecasting based on feature extraction and improved general

regression neural network model," *Energy*, vol. 166, pp. 653–663, 2019.

- [4] G. Cai, X. Liu, C. Dai, L. Zong, and X. Luo, "Load state identification method for wet ball mills based on the meemd singular value entropy and pnn classification," *Mining, Metallurgy & Exploration*, vol. 37, no. 2, pp. 543–553, 2020.
- [5] X. Yan and M. Jia, "Application of csa-vmd and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 122, pp. 56–86, 2019.
- [6] Y. Zhang, B. Chen, G. Pan, and Y. Zhao, "A novel hybrid model based on vmd-wt and pca-bp-rbf neural network for short-term wind speed forecasting," *Energy Conversion and Management*, vol. 195, pp. 180–197, 2019.
- [7] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, "A new prediction method based on vmd-prbf-arma-e model considering wind speed characteristic," *Energy Conversion and Management*, vol. 203, 2020.
- [8] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of vmd for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mechanical Systems and Signal Processing*, vol. 116, pp. 668–692, 2019.
- [9] X. Zhang, W. Tuo, and C. Song, "Application of meemd-arima combining model for annual runoff prediction in the lower yellow river," *Journal of Water and Climate Change*, vol. 11, no. 3, pp. 865–876, 2020.
- [10] L. Yuxing, Y. Li, X. Chen, and J. Yu, "A novel feature extraction method for ship-radiated noise based on variational mode decomposition and multi-scale permutation entropy," *Entropy*, vol. 19, no. 7, 2017.
- [11] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, and W. Shi, "Deep learning based emotion analysis of microblog texts," *Information Fusion*, vol. 64, pp. 1–11, 2020.
- [12] J. Qiu, L. Du, D. Zhang, S. Su, and Z. Tian, "Nei-tte: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2659–2666, 2020.
- [13] C. Luo, Z. Tan, G. Min, J. Gan, W. Shi, and Z. Tian, "A novel web attack detection system for internet of things via ensemble classification," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5810–5818, 2021.
- [14] Y. Hao and C. Tian, "The study and application of a novel hybrid system for air quality early-warning," *Applied Soft Computing*, vol. 74, pp. 729–746, 2019.
- [15] F. He, J. Zhou, Z. kai Feng, G. Liu, and Y. Yang, "A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with bayesian optimization algorithm," *Applied Energy*, vol. 237, pp. 103–116, 2019.
- [16] G. Chitalia, M. Pipattanasomporn, V. Garg, and S. Rahman, "Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks," *Applied Energy*, vol. 278, p. 115410, 2020.
- [17] Y. Dai and P. Zhao, "A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization," *Applied*

- Energy*, vol. 279, p. 115332, 2020.
- [18] C. Li, “Designing a short-term load forecasting model in the urban smart grid system,” *Applied Energy*, vol. 266, p. 114850, 2020.
- [19] Z. Liu, P. Jiang, L. Zhang, and X. Niu, “A combined forecasting model for time series: Application to short-term wind speed forecasting,” *Applied Energy*, vol. 259, p. 114137, 2020.
- [20] N. Somu, G. R. M R, and K. Ramamritham, “A hybrid model for building energy consumption forecasting using long short term memory networks,” *Applied Energy*, vol. 261, p. 114131, 2020.
- [21] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [22] M. Xia, H. Shao, X. Ma, and C. W. de Silva, “A stacked gru-rnn-based approach for predicting renewable energy and electricity load for smart grid operation,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7050–7059, 2021.
- [23] P. K. R. Maddikunta, Q.-V. Pham, P. B, N. Deepa, K. Dev, T. R. Gadekallu, R. Ruby, and M. Liyanage, “Industry 5.0: A survey on enabling technologies and potential applications,” *Journal of Industrial Information Integration*, p. 100257, 2021.