A Novel Data-Driven Method with Decomposition Mechanism Suitable for Different Periods of Electrical Load Forecasting

XIANG WANG1,2,3, ZHANXIA WU2,3, JUNXIONG GE3, ZHANHAO ZHANG4, PUKUN LU5 AND SHUNJIANG WANG6

1Department of Research Center, Beijing Smart-Chip Microelectronics Technology Co., Ltd., Beijing 102200, China
2Department of Research Center, Shenzhen Smart-Chip Microelectronics Technology Co., Ltd., Shenzhen 518048, China
3Department of Research Center, Shenzhen Guodian Technology Communication Co., Ltd., Shenzhen 518000, China
4Department of Research Center, Laijin (Tianjin) Auto Parts Co., Ltd., Tianjin 300000, China
5Department of Automation Engineering, Northeast Electric Power University, Jilin 132012, China
6Department of Dispatch Center, State Grid Liaoning Electrical Power Co., Ltd., Shenyang 110006, China

Corresponding author: JUNXIONG GE (1115927027@qq.com) and PUKUN LU (loopkun@163.com)

This work was supported by the Computer Software Commissioned Development Project of Shenzhen Guodian Technology Communication Co., Ltd. under Grant SGITZXSZWZXZ1201391.

ABSTRACT For improving the precision of the load forecasting in different time spans, a new load forecasting model which combines the improved complete ensemble empirical mode decomposition algorithm based on adaptive noise (ICEEMDAN) algorithm, the least squares support vector machine (LS-SVM) and the long short-term memory network (LSTM) is proposed. In this paper, the training set of the forecasting model is acquired from the department of dispatch center from a large city in the north of China. And the advantages of the forecasting algorithms and decomposition algorithm are applied reasonably, where the ICEEMDAN algorithm is used to decompose the original historical load data. By using the ICEEMDAN algorithm, the fluctuation trend of different periods can be obtained. And the structure of the neural network combining LS-SVM and LSTM, which is used to obtain the load forecasting result. Based on LS-SVM and LSTM algorithm the non-stationary and stationary signals have been processed, respectively. The simulation results show that whether testing by the dataset acquired from a city in the north of China or the Elia dataset, the proposed method outperforms the load forecasting model in the short-, medium- and long-term which is based on PSO-SVR, LS-SVM, LSTM and ICCEMDAN-LSTM, respectively.

INDEX TERMS Load forecasting, Improved complete ensemble empirical mode decomposition algorithm based on adaptive noise, Least squares support vector machine, Long short-term memory network.

I. INTRODUCTION

Recently, load forecasting has played an important role in the future dispatch planning of the power grid, as well as the operation of power market transactions and the economic and stable operation of the power grid system [1,2]. The accuracy of load forecasting directly affects the safety and reliability of the grid [3]. However, with continuous economic development, the load structure of power users has become increasingly complex and the peak-to-valley differences in power grids have continued to increase, causing great difficulties for accurate load forecasting [4]. Therefore, considering the multiple characteristics, strong nonlinearity and strong volatility in load data, it is important to explore how to mine the characteristics’ information, improve the accuracy and efficiency of forecasting, and identify the load forecasting models.

According to the time span, load forecasts are usually classified into three categories: the first is short-term load forecasting (STLF), which is used to forecast load data from a few days. The second category is medium-term load forecasting (MTLF), which is used to forecast load data over a span of weeks. The third category is long-term load forecasting (LTLF), which is used to forecast load data for a month or more [5,6].
The most common load forecasting method is based on time series, which mainly includes statistical models, artificial intelligence (AI) models and hybrid models [7]. Based on only one type of time span, many efforts about load forecasting have been made. Takeda et al. proposed an ensemble Kalman filter (EnKF) method for short-term load forecasting, which improved the load forecasting accuracy [8]. Hassan et al. developed a system for electricity load demand forecasting based on the theory of extreme learning machine (ELM), and the results show that the proposed method was effective [9]. Eskandari et al. used convolutional neural networks for multidimensional feature extraction which were input into a two-way relayed long-term and short-term memory network with gated loop units. They also made it possible to obtain hour-by-hour electricity load forecasting results with the effects of multiple external factors [10]. Samuel et al. proposed an improved entropic mutual information-based feature selection method for data preprocessing. This type of entropy can handle load data of both linear and nonlinear electricity and was used by [11] to remove irrelevant and redundant features. They also combined a conditional restricted Boltzmann machine (CRMB) and a meta heuristic optimization algorithm, Jaya, to construct a medium-term load forecasting model, which improved the algorithm’s convergence speed and reduced execution time. Mohammed and Al-Bazi [12] developed an improved neural network model for long-term load prediction by using the adaptive back propagation algorithm (ABPA), which provides a high prediction accuracy. And the load forecasting methods which have been introduced in the above part are summarized in Table 1.

However, there remains the reason for the plenty of historical load data and many uncertain factors in medium-term and long-term load forecasting. Therefore, the realization of medium-term and long-term load forecasting are the most difficult parts of the study for the load forecasting work. With the improvement in generalization capability of load forecasting algorithms, some load forecasting models can be applied to both types of time span, for example, in [13], combining the weighted least squares state estimation with an adaptive neuro-fuzzy inference system enabled the construction of a hybrid model for short-term load forecasting, which was applicable to medium-term load forecasting. In [14], an artificial neural network model and a stochastic/continuous autoregressive model were developed, which fit the needs of medium-short-term and very short-term load forecasting. In [15], combining the time-based convolutional neural networks and period-based long-short-term memory networks made it possible to identify load forecasting models for the medium- and short-term forecasting, which reduced the computational complexity. In [16], a kind of hierarchical optimization method was proposed, which combines a nested-based strategy, state transfer algorithm and hybrid support vector regression, thereby fitting the needs of medium and long-term load forecasting.

In summary, many researchers have developed helpful load forecasting algorithms which are applicable to one or two types of time span. However, according to our recent search results, it is still a challenge to build load forecasting models which are applicable to short-term, medium-term and long-term at the same time. To solve this problem, the authors of this paper have developed a forecasting model which combines an improved complete ensemble empirical mode decomposition algorithm based on adaptive noise and weighted least squares-based support vector machine algorithm, as well as a long-short-term memory neural network. To best of our knowledge, this is the first time the forecasting algorithm, which combines the improved complete ensemble empirical mode decomposition algorithm based on adaptive noise (ICEEMDAN) model, as well as LS-SVM and LSTM has been applied to load forecasting research. By using the actual power load data and holiday data, the generalization ability of load forecasting in the short-term, medium-term and long-term can be verified. The contributions of this model are as follows:

1) This paper proposes a novel hybrid model based on

| Time Span        | Forecasting Methods          | Features Extraction | Forecasting Algorithm | Reference |
|------------------|------------------------------|---------------------|-----------------------|-----------|
| short-term       | statistical models           | -                   | EnKF, MLR, lasso      | [8]       |
| short-term       | artificial intelligence (AI) models | -                   | ELM                   | [9]       |
| short-term       | hybrid models                | CNN                 | LSTM, GRU             | [10]      |
| medium-term      | hybrid models                | Adaptive K-means    | CRMB, Jaya            | [11]      |
ICEEMDAN, LS-SVM and LSTM. By using ICEEMDAN, the load data can be decomposed into stationary and non-stationary signals, and the neural network input data processing can be completed. In the structure of the neural network combining LS-SVM and LSTM, the LS-SVM network is used to process the input non-stationary signal and the LSTM network is used to process the input stationary signal.

2) The hybrid model is applied to short-term, medium-term and long-term load forecasting. In this paper, the advantages of ICEEMDAN, LS-SVM and LSTM are comprehensively utilized. Based on the ICEEMDAN algorithm, the fluctuation trend of different time spans in load data series can be accurately extracted. By using LS-SVM and LSTM to give full play to their respective technical advantages of coping with non-stationary and stationary signals, which can accurately learn the fluctuation characteristics of different time spans in each sub-series, and then realize the short-term, medium-term and long-term load forecasting.

3) The historical load data measurement at 1-minute intervals from a city of northern China and the holiday data is chosen as the training set, which improves the accuracy of the load forecasting model. The model is trained and tested by the actual data set, and the accuracy of the proposed model is verified to be improved compared to the existing models for load forecasting in short-term, medium-term and long-term.

II. FORECASTING ARCHITECTURE

The original power load data is an effective basis for power load forecasting. However, fluctuation and uncertainty in the data can have a great impact on forecasting results, even causing the predicted values to deviate greatly from the real results. Therefore, in this paper, a method is developed using an improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) will be adopted to preprocess power load data, and the signals of data will be decomposed into stationary and non-stationary. By using the corresponding neural network, the strong generalization ability of the proposed forecasting scheme can be realized.

A. COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITION WITH ADAPTIVE NOISE (ICEEMDAN)

ICEEMDAN is an advanced theoretical achievement based on empirical mode decomposition (EMD) [17]. The traditional EMD method can be used to analyze the nonlinear signal sequence, and the natural mode function is adaptively generated by the analyzed signal, so as to realize the decomposition of the natural mode function and residual in the signal [18]. Based on the EMD algorithm, the mode aliasing can be effectively restrained by adding limited white noise. [19] However, with the increase of white noise frequency, the efficiency of the algorithm can gradually worsen, and the prediction error can increase after reconstruction. To deal with this problem, CEEMDAN transforms the white noise into adaptive white noise, which can make the signal reconstruction error approach zero. This method can adaptively decompose nonlinear and non-stationary signals into multiple natural modal components with different scales. Let \( I(t) \) serve as the original signal sequence. \( \xi_k, k = 0, ..., K \) are the adaptive coefficient, and \( w(t) \) are the white noise sequence added each time, which are selected as the realization of zero mean unit variance white noise here. \( E_t(\cdot) \) is the \( k \) th modal operator generated by EMD. Let \( \text{IMF}_t(t) = I(t) + \xi_k w(t) \) \( \text{IMF}_t(t) \) as the first natural modal component, \( R_t(t) \) as the first remainder sequence, and the CEEMDAN algorithm is shown in algorithm 1.

Algorithm I: Complete ensemble empirical mode decomposition with adaptive noise

1: Use empirical mode decomposition (EMD) to decompose N times
   \( \text{IMF}_t(t) = \frac{1}{N} \sum_{i=1}^{N} \text{IMF}_{t,i}(t) \)
2: Calculate the first remainder sequence
   \( R_t(t) = I(t) - \text{IMF}_t(t) \)
3: Calculate the second natural modal component:
   \( \text{IMF}_{t+1}(t) = \frac{1}{N} \sum_{i=1}^{N} E_i(R_t(t) + \xi_i E_t(w(t))) \)
4: for k=2 to K do
5: if extreme point\(>2 \) then
6:   \( R_t(t) = R_{t-1}(t) - \text{IMF}_{t+1} \)
   \( \text{IMF}_{t+1}(t) = \frac{1}{N} \sum_{i=1}^{N} E_i(R_t(t) + \xi_i E_t(w(t))) \)
7: end if
8: end for
9: End of the decomposition, obtain
   \( I(t) = \sum_{i=1}^{N} \text{IMF}_{t,i}(t) + R(t) \)

However, the CEEMDAN algorithm will produce pseudo modes in the early stage of decomposition, and the residual noise, which degrades the performance of modal decomposition still exists [20].
In order to eliminate the pseudo modes generated by CEEMDAN and improve the performance of the modal decomposition algorithm, the improved complete ensemble empirical mode decomposition algorithm based on adaptive noise (ICEEMDAN) method is proposed. The ICEEMDAN algorithm was proposed by Comomainas based on the CEEMDAN algorithm, which was applied in much research [21,22]. Unlike CEEMDAN, ICEEMDAN uses the local average achieved by additive white noise estimation instead of the white noise estimation mode. Define \( M(\cdot) \) as the envelope operator; let \( E_i(I) = I - M(I) \), and define \( w_i(t) \) as the realization of Gaussian white noise with zero mean and unit variance. Then, the detailed steps of ICEEMDAN are shown as follows.

Step 1: Owing to \( I_i(t) = I(t) + \xi_i E_i(w_i(t)), \ i = 1, 2, \ldots, I \), the first residual sequence can be obtained as:

\[
R_i = \frac{1}{I} \sum_{i=1}^{I} M(I_i(t))
\]  

(1)

Step 2: Define the first mode as:

\[
IMF_1 = I(t) - R_i
\]  

(2)

Step 3: Calculate the second residual sequence \( R_2 \), and the second model \( IMF_2 \) can be obtained as:

\[
R_2 = \frac{1}{I} \sum_{i=1}^{I} M(R_i + \xi_2 E_i(w_i(t)))
\]  

(3)

\[
IMF_2 = R_i - R_2
\]

Step 4: Calculate the \( k \)th residual sequence \( R_k \), and the second model \( IMF_k \) can be obtained

\[
R_k = \frac{1}{I} \sum_{i=1}^{I} M(R_{k-1} + \xi_k E_i(w_i(t)))
\]  

(4)

\[
IMF_k = R_{k-1} - R_k
\]

Step 5: Repeat step 4 until all the intrinsic modal components are obtained.

C. LONG SHORT-TERM MEMORY NETWORK (LSTM)

The LSTM is an extension of the recurrent neural network (RNN) and has been improved compared to the original RNN network. The main difference between RNNs and LSTMs is that LSTMs can store long-term time-dependent information and can appropriately map input and output data, providing a strong capability in predicting time-series data [23].

The LSTM network structure is added with cells containing the control information flow and three gates in comparison with the traditional perceptron architecture. These three gates are the input, forget, and output gates [24]. The three gates have different roles and cooperate with each other to achieve the best results. The LSTM uses the three gates to control the retained information and ensure that the retained information is what the algorithm needs. The structure diagram of the LSTM is shown in Fig. 2.

![FIGURE 2. Structure of the LSTM.](image)

where \( x(t) \) is the input; \( h(t_{i-1}) \) and \( h(t) \) are the output values at times \( t_{i-1} \) and \( t_i \), respectively. \( c(t_{i-1}) \) and \( c(t) \) are the cell states at times \( t_{i-1} \) and \( t_i \); \( b = \{b_x, b_f, b_o, b_c \} \) are the deviation of the input gate, the forgetting gate, the internal state and the output gate, respectively. \( W_i = \{w_x, w_f, w_o, w_c \} \) are the weight matrix of the input gate, the forgetting gate, the internal state and the output gate, respectively. \( W_2 = \{w_{ix}, w_{if}, w_{io}, w_{ic} \} \) is the cycle weight; \( a = \{a(t), f(t), c(t), o(t) \} \) are the output result of the input gate, the forget gate, the internal state and the output gate, respectively.

The LSTM operates according to the following steps [25]:

1. The primary task of the data entering the LSTM is to be detected by the forget gate, which focuses on determining what information needs to be removed from the data, that is, what information can be dropped from the cells. This

FIGURE 1. The structure block diagram of ICEEMDAN.

It should be noted that the selection of constant \( \xi_{k-1} \) can change the signal-to-noise ratio of the residual and noise, and satisfying \( \xi_0 = \varepsilon_o \cdot \text{std}(I) / \text{std}(E_i(w_i)) \) and \( \xi_k = \varepsilon_i \cdot \text{std}(R_i) \), with \( \varepsilon_i \) being the reciprocal of the expected signal-to-noise ratio of the input signal \( I(t) \) to the first noise, and \( \text{std}(\cdot) \) being the standard deviation operator. The structure block diagram of ICEEMDAN is shown in Fig. 1.
strategy is implemented through a sigmoid layer called the forget gate.
\[ f(t) = \sigma(w_x x(t) + w_h h(t_{i-1}) + b) \]  
(5)

(2) The second step of the LSTM is to determine which new information will be saved within the cell. First, the sigmoid layer of the input gates is used to decide which values are in the update cell.
\[ a(t) = \sigma(w_x x(t) + w_h h(t_{i-1}) + b) \]  
(6)

Second, the tanh layer is used to create new values to be added to the cell. This means the results of the forget gates and input gates must be combined to create a state update that updates the old cell state \( c(t_{i-1}) \) to the new cell state \( c(t) \). 
\[ c(t) = f_t \times c(t_{i-1}) + a_t \times \tanh(w_x x(t) + w_h h(t_{i-1})) + b \]  
(7)

(3) The final step of the LSTM is to identify the results of the cell state outputs. The strategy is implemented in two parts.
First, a sigmoid layer is used to determine which parts of the cell state are output.
\[ o(t) = \sigma(w_x x(t) + w_h h(t_{i-1}) + b) \]  
(8)

Then, the output of the cell is set to the tanh functions multiplied by the output of the sigmoid layer so that only the part needed is the output.
\[ h(t) = o(t) \times \tanh(c(t)) \]  
(9)

where tanh is the tangent excitation function and \( \times \) denotes the point-by-point multiplication.

D. LEAST SQUARES SUPPORT VECTOR MACHINE (LS-SVM)

Uyken et al. proposed the least squares support vector machine (LS-SVM) [26-28] in recent years, which is one of the most important achievements in statistical theory. The LS-SVM training process follows the principle of structural risk minimization and has the advantages of low computational complexity and high operational speed compared with general vector machines.

The SVM employs a linear model to predict sample data by using a nonlinear mapping from the input vector to a high-dimensional feature space. Considering the given training set \( X = \{ (x_i, y_i) \}_{i=1,2,...,m} \), input data \( x_i \in \mathbb{R}^n \), and output data \( y_i \in \mathbb{R} \), the LS-SVM aims to find a plane \( w^T x + b = 0 \) to separate the two classes of samples, that is, the solution of the following quadratic programming problem [29], which is calculated as:
\[ \min \frac{1}{2} \| w \|^2 + \frac{1}{2} \sum_{i=1}^l \xi_i^2 \]  
(10)
\[ s.t. y_i [w \cdot \phi(x) + b] = 1 - \xi_i^2, i = 1, 2, \ldots, l \]  
(11)
where \( \gamma \) denotes the penalty parameter and \( \xi \) denotes the slack variable.

For (10), the Lagrangian operator can be introduced to obtain the Lagrangian objective function, which is expressed as:
\[ L(w, b, \xi, \alpha_i) = \frac{1}{2} \| w \|^2 + \frac{1}{2} \sum_{i=1}^l \xi_i^2 \]  
(12)
\[ - \sum_{i=1}^l \alpha_i [y_i [w \cdot \phi(x) + b] - 1 + \xi_i]. \]

where \( \alpha_i \) is the Lagrangian multiplier. Let the partial derivatives of each parameter of (12) be zero; then, the Karush-Kuhn-Tucker (KKT) optimality condition for the solution can be obtained as:
\[ \frac{\partial L}{\partial \xi} = 0 \Rightarrow \alpha_i = y_i [w \cdot \phi(x) + b] - 1 + \xi_i = 0. \]  
(13)

(13) can then be expressed as a system of equations:
\[ \begin{bmatrix} \Omega & \gamma & \alpha \end{bmatrix} \begin{bmatrix} y \ 0 \ b \end{bmatrix} = \begin{bmatrix} e \ 0 \ end{bmatrix} \]  
(14)

where \( \Omega_i = y_i y_i K(x, x_i) \).

Solving for \( \alpha \) and \( b \) according to (14), the classification discriminant function is obtained as:
\[ y(x) = \text{sign} \left( \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \right) \]  
(15)

There are several different types of Mercer kernel functions \( K(x, x_i) \), such as sigmoid, polynomial, and radial basis functions (RBFs). RBF is a common option for kernel functions because of the fewer parameters that need to be set and the excellent overall performance [30]. Therefore, in this paper, RBF is chosen as the kernel function, which is given as:
\[ K(x, x_i) = \exp \left\{ -\| x - x_i \|^2 / 2\sigma^2 \right\} \]  
(16)

E. The Proposed Method

When power grid planning, different effects will be produced while in different terms (time periods) of load forecasting. Short-term load forecasting (STLF) is mostly applied to monitor the power quality of the grid. Medium term load forecasting (MTLF) is mainly used for making generation plans and guiding electric energy production, and long-term load forecasting (LTFL) is used in power grid construction planning. In this article, we propose a hybrid algorithm that includes ICEEMDAN, LS-SVM, and LSTM for solving the load forecasting problems of short, medium, and long terms. Here, ICEEMDAN is used to decompose the original load data onto the stationary signals and non-stationary signals. Then, we put the signals into the forecasting algorithms and get the result.

Remark 1: In this paper, the dataset was acquired from a city in the north of China. And the results of the load forecasting
can apply to guiding electric energy production and provide auxiliary decision-making for the actual power system.

The flow chart of the proposed hybrid algorithm is shown in Fig. 3.

The steps of the proposed hybrid algorithm are shown as follows:

Step 1: Data Analysis
The purpose of this step is to ensure the quality of the dataset when the historical load data must be cleaned. First, the missing values and outliers must be checked to avoid the effects of missing values and outliers in the dataset. The missing values and outliers are repaired by the trend of the historical load and median value. In this paper, we use the linear interpolation to deal with this issue, which can be expressed as (17)-(19).

\[
L(x) = y_i(x) + y_i'(x) \quad (17)
\]
\[
l_i(x) = \frac{y_i(x) - y_i(x)}{y_i(x) - y_i(x)} \quad (18)
\]
\[
l_i(x) = \frac{y_i(x) - y_i(x)}{y_i(x) - y_i(x)} \quad (19)
\]

where the interpolation formula \( L(x) \) is approximated by a line that goes through \((x_0, y_0) \) and \((x_i, y_i) \).

Step 2: Data Decomposition
In Step 2, the load data will be decomposed by ICEEMDAN. According to the ICEEMDAN principle, the load data will be decomposed into IMF, and a margin RES, by k-step iteration, where IMF and RES are the stationary signal and the non-stationary signal which obtained by decomposition, respectively. Then, to train and evaluate the forecasting model established in step 3, the load data is divided into training set and test set.

Step 3: Load Forecasting
In this step, we put two parts of the load data decomposed by ICEEMDAN into the forecasting algorithm. This requires taking the non-stationary signal and the stationary signal into the forecasting algorithm as input, respectively. In order to establish the load prediction model and adjust the model parameters, the non-stationary signal is brought into the LS-SVM algorithm, and the stationary signal is brought into the LSTM algorithm. Then, we sum the output of the two algorithms to get the result of forecasting. Finally, the load prediction results are obtained by normalizing, which restores the original dimension of the data.

Step 4: Model Validation
In Step 4, we verify the accuracy of the model. The degree of fitting and accuracy of the established model is evaluated by the evaluation indices. The evaluation indices are introduced in the next section.

FIGURE 3. Flow chart of the proposed hybrid algorithm.
III. RESULTS AND DISCUSSION
To realize the proposed algorithm, we chose the historical load data measured at 1-minute intervals, is utilized and resampled at 15-minute intervals. Then, we chose the holiday data from 2018 to 2020 from a city of northern China to build a load forecasting model which is written in Python. In this model, the data set is divided into training set and test set in 3:1. Then the PSO-SVR-based, LS-SVM-based, LSTM-based and ICEEMDAN-LSTM-based forecasting method are compared and the results of the simulation are obtained. It should be noted that the electricity consumption in summer and winter is more than that in spring and autumn because the air conditioners are often used for cooling and heating in summer and winter, respectively. Usually, the last IMF component contains the seasonal trend of the load data. And the seasonal trend can be captured and learned by the LSTM algorithm effectively, which provides important information for long-term load forecasting. Here the result of the decomposition by four months and the result of the decomposition by a year are shown in Fig.4 and Fig.5.

Remark 2: In order to construct the load forecasting model, the real load data from a large industrial city in the north of China was obtained in this research to construct and test the load forecasting model. Thus, the real-world scenario in this research is a city in the north of China.

| Algorithm      | Type       | Parameter | Type       | Parameter |
|----------------|------------|-----------|------------|-----------|
| LSTM           | Layers     | 2         | Epochs     | 10        |
|                | Uints      | (50, 50)  | Bach Size  | 128       |
| LS-SVM         | Kernel Function | RBF    | Loss Factor (ε) | 1         |
|                | Penalty Factor (C) | 0.6     | Kernel Factor (γ) | 10        |
| PSO-SVR        | n_iterations | 50      | Learning Factor (C1) | 2         |
|                | inertia_weight | 0.5     | Learning Factor (C2) | 2         |
| OUR PROPOSED   | Layers     | 2         | Epochs     | 10        |
|                | Uints      | (50, 50)  | Bach Size  | 128       |
|                | Kernel Function | RBF    | Loss Factor (ε) | 1         |
|                | Penalty Factor (C) | 0.6     | Kernel Factor (γ) | 10        |

FIGURE 4. The result of the decomposition by four months.
In the process of simulation, the model parameters for the load forecasting in the three time spans we selected are shown in Table II.

At the end, to evaluate the accuracy of the proposed algorithm, three metrics were selected: mean average error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). The expressions of these three evaluation criteria are:

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |C_i - R_i| 
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{R_i - C_i}{R_i} \right| 
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (C_i - R_i)^2} 
\]

where \(N\) is the number of the data point, \(C_i\) is the result of the forecast data and \(R_i\) is the real data (\(i = 1 \cdots N\)).

By using the load forecasting model and the three metrics. The forecasting results of the short term, the medium term and the long term will be discussed in next three subsections, respectively.

A. SHORT-TERM LOAD FORECASTING (STLF)

To forecast the electricity load of the last three days of 2020, the effectiveness of the proposed hybrid algorithm in STLF was validated, and the training set was three times as long as the test set. The STLF based on the proposed method and its comparison with the PSO-SVR, LS-SVM, LSTM and ICCEMDAN-LSTM are shown in Fig. 6.

The correlation curve between actual value and predicted value is shown in Fig. 7. And the MAE, MAPE, and RMSE of SLTF are listed in Table III.

Table III shows the prediction effect of using only LSTM or LS-SVM to be poor. The prediction results are significantly improved by mixing ICEEMDAN, LSTM and LS-SVM.
FIGURE 7. The correlation curve between actual value and predicted value in STLF.

The proposed algorithm provides a mean average error (MAE), mean average percentage error (MAPE) and the root mean squared error (RMSE) of 334.8162, 0.6049% and 371.0061 respectively in STLF which provides 949.3582, 542.1562, 284.0131 and 159.8592 less MAE compared to PSO-SVR, LS-SVM, LSTM and ICEEMDAN-LSTM, respectively.

| Algorithm          | Evaluation Criteria | MAE    | MAPE    |
|--------------------|---------------------|--------|---------|
|                    |                     | PSO-SVR| 1284.1744| 2.2497% |
|                    |                     | LS-SVM | 876.9724 | 1.6527% |
|                    |                     | LSTM   | 618.8293 | 1.1812% |
|                    |                     | ICEEMDAN-LSTM | 494.6754 | 0.9340% |
| OUR PROPOSED       |                     |        | 334.8162 | 0.6049% |

FIGURE 8. Comparison of the proposed method with PSO-SVR, LS-SVM, LSTM and ICEEMDAN-LSTM in MTLF.

FIGURE 9. The correlation curve between actual value and predicted value in MTLF.

TABLE III
THE PERFORMANCE OF STLF

| Algorithm          | Evaluation Criteria | MAE    | MAPE    | RMSE    |
|--------------------|---------------------|--------|---------|---------|
| PSO-SVR            |                     | 1263.5849 | 2.3136% | 1363.6699 |
| LS-SVM             |                     | 761.2472 | 1.4796% | 778.4756 |
| LSTM               |                     | 541.9591 | 0.9429% | 647.1733 |
| ICEEMDAN-LSTM      |                     | 409.1145 | 0.8270% | 457.5631 |
| OUR PROPOSED       |                     | 426.1323 | 0.7829% | 455.2315 |

B. MEDIUM-TERM LOAD FORECASTING (MTLF)

The electricity load data compiled during the last 4 weeks of 2020 was utilized for MTLF forecasting, with the training and test datasets divided in a ratio of 3:1. The forecasting result of MTLF and the correlation curve between the actual value and predicted value are shown in Fig. 8. and Fig. 9, respectively. The MAE, MAPE, and RMSE of MTLF are listed in Table IV.

TABLE IV
THE PERFORMANCE OF MTLF

| Algorithm          | Evaluation Criteria | MAE    | MAPE    |
|--------------------|---------------------|--------|---------|
|                    |                     | PSO-SVR| 1263.5849 | 2.3136% |
|                    |                     | LS-SVM | 761.2472 | 1.4796% |
|                    |                     | LSTM   | 541.9591 | 0.9429% |
|                    |                     | ICEEMDAN-LSTM | 409.1145 | 0.8270% |
| OUR PROPOSED       |                     |        | 426.1323 | 0.7829% |

As shown in Table IV, with the increase of prediction sequence, the prediction performance of the proposed algorithm weakens slightly. However, our proposed hybrid algorithm is obviously superior to other algorithms. The MAPE of the proposed method was reduced by 1.5307%, 0.6967%, 0.16% and 0.0441% respectively, compared with the PSO-SVR, LS-SVM, LSTM and ICEEMDAN-LSTM.
C. LONG-TERM LOAD FORECASTING (LTLF)

FIGURE 10. Comparison of the proposed method with PSO-SVR, LS-SVM, LSTM and ICEEMDAN-LSTM in LTLF.

FIGURE 11. The correlation curve between the actual value and predicted value in LTLF.

Following the same protocol as in the above two cases, the power load data from September to November was used to train the data-driven model and to forecast the load of December before the model validity for LTLF could be verified. The simulation results are shown in Fig. 10 and Fig. 11, where Fig. 10 is the result of the LTLF, and Fig. 11 is the correlation curve between the actual value and predicted value.

The MAE, MAPE, and RMSE of LTLF are listed in Table V, and the comparison of MAPE by using experimental sets in STLF, MTLF and LTLF is shown in Fig. 12.

From the results of Table V, the proposed method reflects the effectiveness. Clearly, the RMSE was reduced by 855.7926, 505.3772, 226.5503 and 23.9541, respectively, when comparing with the PSO-SVR, the LS-SVM, LSTM and the ICEEMDAN-LSTM. According to Fig. 5, Fig. 7 and Fig. 9, it can be seen that the fitting effect of LTLF is the best because the correlation curve between actual value and predicted value in LTLF has the largest $R^2$.

Remark 3: As shown in Table III, Table IV and Table V, through comparing the ICEEMDAN-LSTM model with the LSTM model, it is obvious that the accuracy of the load forecasting has been improved by using the ICEEMDAN algorithm. Comparing the ICEEMDAN-LSTM model with our proposed model, we can conclude that the simulation results of the LS-SVM-LSTM model, which cope with the non-stationary and stationary signals in a model, respectively, can enhance the performance of the load forecasting.

D. MODEL VALIDATION OF AN OPEN SOURCE POWER LOAD DATASET

To verify the generalization ability of the proposed hybrid algorithm, the electricity load data collected from the European Elia Grid [31] was utilized to perform load forecasting, including the measured net generation of the (local) power stations that inject power into the Elia grid. The netto inflows from the Distribution to the Elia Grid and the netto import on borders. Exports on borders and energy used for energy storage are deducted.

Then, we also selected the PSO-SVR, LS-SVM, LSTM and ICEEMDAN-LSTM for comparison. It should be noted that the model parameters for the load forecasting in three types of spans were conducted by using a common experimental set with the same parameters as found in Table II.

![Image](image-url)
FIGURE 13. The results of STLF in a common experimental setup.

FIGURE 14. The result of MTLF in a common experimental setup.

FIGURE 15. The result of LTLF in a common experimental setup.

FIGURE 16. The correlation curve between the actual value and predicted value in STLF were determined by using a common experimental setup.

FIGURE 17. The correlation curve between the actual value and predicted value in MTLF was obtained in a common experimental setup.

FIGURE 18. The correlation curve between actual value and predicted value in LTLF by using a common experimental setup.
FIGURE 19. Comparison of MAPE by using common experimental setups in STLF, MTLF and LTLF.

By using this common experimental setup, we obtained the results of STLF, MTLF and LTLF in Fig. 13, Fig. 14, and Fig. 15 respectively. The correlation curve between the actual value and predicted value in the three types of time spans are shown in Fig. 16, Fig. 17, and Fig. 18 respectively.

The MAE, MAPE, and RMSE of STLF, MTLF and LTLF were obtained by using common experimental setup as listed in Table VI and a comparison of MAPE can be made by using the experimental setup in STLF, MTLF and LTLF as shown in Fig. 19. Table VI indicates that the proposed method performed well in all time spans by using the common experimental set. In all cases, compared with PSO-SVR, LS-SVM, LSTM and ICCEMDAN-LSTM, our method had the smallest MAE, MAPE and RMSE. It also adds to the proof that an algorithm has good robustness and generalization ability.

| Type   | Metrics | Algorithm   | MAE      | Algorithm   | MAPE     | Algorithm   | RMSE      |
|--------|---------|-------------|----------|-------------|----------|-------------|-----------|
| STLF   |         | PSO-SVR     | 201.5045 | PSO-SVR     | 2.3760%  | PSO-SVR     | 248.7796  |
|        |         | LS-SVM      | 178.8277 | LS-SVM      | 2.0428%  | LS-SVM      | 218.7950  |
|        |         | LSTM        | 159.6796 | LSTM        | 1.8426%  | LSTM        | 190.7682  |
|        |         | ICEEMDAN-LSTM | 137.0839 | ICEEMDAN-LSTM | 1.5920% | ICEEMDAN-LSTM | 165.1453  |
|        |         | Our Proposed | 105.5437 | Our Proposed | 1.2181% | Our Proposed | 126.3512  |
| MTLF   |         | PSO-SVR     | 197.1316 | PSO-SVR     | 2.5141%  | PSO-SVR     | 238.5954  |
|        |         | LS-SVM      | 162.0518 | LS-SVM      | 2.0231%  | LS-SVM      | 194.165   |
|        |         | LSTM        | 158.7912 | LSTM        | 1.9485%  | LSTM        | 192.9195  |
|        |         | ICEEMDAN-LSTM | 124.0841 | ICEEMDAN-LSTM | 1.5819% | ICEEMDAN-LSTM | 150.5812  |
|        |         | Our Proposed | 109.9484 | Our Proposed | 1.3399% | Our Proposed | 133.3694  |
| LTLF   |         | PSO-SVR     | 199.2078 | PSO-SVR     | 2.2944%  | PSO-SVR     | 244.2685  |
|        |         | LS-SVM      | 181.7951 | LS-SVM      | 2.0743%  | LS-SVM      | 221.2519  |
|        |         | LSTM        | 176.6144 | LSTM        | 1.9552%  | LSTM        | 221.5527  |
|        |         | ICEEMDAN-LSTM | 135.2650 | ICEEMDAN-LSTM | 1.5233% | ICEEMDAN-LSTM | 165.2524  |
|        |         | Our Proposed | 115.1949 | Our Proposed | 1.2836% | Our Proposed | 142.2339  |

IV. CONCLUSION

This paper presents a forecasting method combining ICEEMDAN, LS-SVM and LSTM, which the authors developed to forecast the electric load for three types of time spans: short, medium and long term. The proposed method is implemented by using the hidden features of the least squares support vector machine (LS-SVM) and LSTM networks to obtain the advantages of both algorithms. The performance of the proposed algorithm was evaluated by investigating real data of power loads in the power system of a large city in China. To ensure stability and validity, the validation corresponds to the use of a dataset from each section. In addition, the robustness and generality of the designed scheme were tested. From the experimental part, we observed that the proposed algorithm gives the lowest MAE, RMSE and MAPE values compared to PSO-SVR, LS-SVM, LSTM and ICCEMDAN-LSTM algorithms. In all validation
experiments, the proposed scheme outperforms the PSO-SVR, LS-SVM, LSTM and ICCEMDAN-LSTM algorithms in terms of prediction results. Finally, it can be shown that the proposed scheme can handle long time series electric load data and predict future loads for a considerable period of time. In addition, it can also handle load forecasting for three types of time horizons simultaneously: short, medium and long term.

REFERENCES

[1] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu and J. He, “Short-Term load forecasting with deep residual networks,” IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[2] L. Cai, J. Gu and Z. Jin, “Two-layer transfer-learning-based architecture for short-term load forecasting,” IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 1722–1732, Mar. 2020.

[3] Z. Yu, Z. Niu, W. Tang and Q. Wu, “Deep learning for daily peak load forecasting—A novel gated recurrent neural network combining dynamic time warping,” IEEE Access, vol. 7, pp. 17184–17194, Jan. 2019.

[4] H. S. Hippert, C. E. Pedreira and R. C. Souza, “Neural networks for short-term load forecasting: a review and evaluation,” IEEE Transactions on Power Systems, vol. 16, no. 1, pp. 44–55, Feb. 2001.

[5] K. B. Debnath and M. Moursheid, “Forecasting methods in energy planning models,” Renewable and Sustainable Energy Reviews, vol. 88, pp. 297-325, May 2018.

[6] K. G. Borojeni, M. H. Amini, S. Bahrami, S. S. Lyengar and A. I. Sarwat, “A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon,” Electric Power Systems Research, vol. 142, pp. 58-73, Jan. 2017.

[7] D. Wang, H. Luo, O. Grunder, Y. Lin and H. Guo, “Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm,” Applied Energy, vol. 190, no. 15, pp. 390–407, Mar. 2017.

[8] H. Takeda, Y. Tamura and S. Sato, “Using the ensemble Kalman filter for electricity load forecasting and analysis,” Energy, vol. 104, no. 1, pp. 184–198, Jun. 2016.

[9] S. Hassan, A. Khosravi, J. Jaafari and M. A. Khanesar, “A systematic design of interval type-2 fuzzy logic system using extreme learning machine for electricity load demand forecasting,” International Journal of Electrical Power & Energy Systems, vol. 82, pp. 1–10, Nov. 2016.

[10] H. Eskandari, M. Imani and M. P. Moghaddam, “Convolutional and recurrent neural network based model for short-term load forecasting,” Electric Power Systems Research, vol. 150, pp. 105713, Jun. 2017.

[11] J. O. Samuel, F.A. Alzahrani, R. J. Li, H. Xian, H. Farooq, M. Shaﬁq, M. K. Afzal and N. Javaid, “Towards modiﬁed entropy mutual information feature selection to forecast medium-term load using a deep learning model in smart homes,” Entropy, vol. 33, no. 1, 68, Jan. 2020.

[12] N. A. Mohammed and A. B. Ammar, “An adaptive backpropagation algorithm for long-term electricity load forecasting,” Neural Computing and Applications, pp. 1–15, Aug. 2021.

[13] M. Ali, M. Adnan, M. Tariq and H. V. Poor, “Load forecasting through estimated parametrized based fuzzy inference system in smart grids,” IEEE Transactions on Fuzzy Systems, vol. 29, no. 1, pp. 156–165, Jan. 2021.

[14] K. Moustris, K. A. Kavadias, D. Zafirakis and K. I. Kaldeis, “Medium, short and very short-term prognosis of load demand for the Greek Island of Tinos using artiﬁcial neural networks and human thermal comfort–discomfort biometeorological data,” Renewable Energy, vol. 147, part 1, pp. 100–109, Mar. 2020.

[15] L. Han, Y. Peng, Y. Li, B. Yong, Q. Zhou and L. Shu, “Enhanced deep networks for short-term and medium-term load forecasting,” IEEE Access, vol. 7, pp. 4045–4055, Dec. 2019.

[16] Z. Wang, X. Zhou, J. Tian and T. Huang, “Hierarchical parameter optimization based support vector regression for power load forecasting,” Sustainable Cities and Society, vol. 71, 102937, Aug. 2021.

[17] Y. Li and L. Wang, “A novel noise reduction technique for underwater acoustic signals based on complete ensemble empirical mode decomposition with adaptive noise, minimum mean square variance criterion and least mean square adaptive filter,” Defence Technology, vol. 16, no. 1, pp. 543–554, Jun. 2020.

[18] Y. Sun, Y. Cao, M. Zhou, T. Wen, P. Li and C. Roberts, “A hybrid method for life prediction of railway relays based on multi-layer decomposition and RBFNN,” IEEE Access, vol. 7, pp. 44761–44770, Apr. 2019.

[19] Y. Li, B. Tang, X. Jiang and Y. Yi, “Bearing fault feature extraction method based on GA-VMD and center frequency,” Mathematical Problems in Engineering, vol. 2022, Jan. 2022.

[20] M. A. Colominas, G. Schlotthauer and M. E. Torres, “Improved complete ensemble EMD: A suitable tool for biomedical signal processing,” Biomedical Signal Process & Control, vol. 14, pp. 19–29, Nov. 2014.

[21] M. Torres, M. Colominas, G. Schlotthauer and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in Proc. ICASSP, Prague, Czech Republic, 2011, pp. 4144–4147.

[22] M. L. Boulahsia and M. Noutoua, “The analysis of tool vibration signals by spectral kurtosis and ICCEMDAN modes energy for insert wear monitoring in turning operation,” The International Journal of Advanced Manufacturing Technology, vol. 115, no. 9-10, pp. 2989–3001, Jun. 2021.

[23] P. Coulibaly, F. Ancitl, P. Rasmussen and B. Bobée, “A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff,” Hydrological Processes, vol. 14, no. 15, pp. 2755–2777, Oct. 2000.

[24] X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You and T. Chi, “Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation,” Environmental Pollution, vol. 231, part 1, pp. 997–1004, Dec. 2017.

[25] G. Alex and S. Jürgen, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” Neural Networks, vol. 18, issue 5–6, pp. 602–610, Jul.–Aug. 2005.

[26] V. N. Vapnik, “An overview of statistical learning theory,” IEEE Trans. on Neural Networks, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[27] J. S. Nan, L. Sun, B. Chen, Z. Lin and K. Toh, “Density-dependent quantized least squares support vector machine for large data sets,” IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 1, pp. 94–106, Jan. 2017.

[28] S. Zhang, S. Zhao, Y. Sui and L. Zhang, “Single object tracking with fuzzy least squares support vector machine,” IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5723–5738, Dec. 2015.

[29] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” Neural Processing, Lett. 1999, 3(3): 293-300, vol. 9, pp. 293–300, Jun. 1999.

[30] S. K. Keerthi and C. J. Lin, “Asymptotic behaviors of support vector machine classifiers,” J. Mach. Learn. Res. 2001, vol. 2, pp. 149-185.

[31] V. N. Vapnik, “An overview of statistical learning theory,” IEEE Trans. on Neural Networks, vol. 10, no. 5, pp. 988–999, Sep. 1999.

[32] J. S. Nan, L. Sun, B. Chen, Z. Lin and K. Toh, “Density-dependent quantized least squares support vector machine for large data sets,” IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 1, pp. 94–106, Jan. 2017.

[33] S. Zhang, S. Zhao, Y. Sui and L. Zhang, “Single object tracking with fuzzy least squares support vector machine,” IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5723–5738, Dec. 2015.

[34] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” Neural Processing, Lett. 1999, 3(3): 293-300, vol. 9, pp. 293–300, Jun. 1999.

[35] S. K. Keerthi and C. J. Lin, “Asymptotic behaviors of support vector machines with Gaussian kernel,” Neural Comput., vol. 15, no. 7, pp. 1667–1689, Jul. 2003.

[36] ELIA. The Elia-grid load data (2021, Dec 24). (https://www.elia.be/en/grid-data/data-download-page)
Xiang Wang received the B.S. degrees in international economy and trade from Peking University, Beijing, China, in 2003 and the M.S. in business administration from Beijing Institute of Technology, Beijing, China, in 2009. He is currently an intermediate economist in Beijing Smart-Chip Microelectronics Technology Co., Ltd.. His current research interests include electricity market analysis and power system load forecasting.

Zhanxia Wu received the M.S. in mechanical practice and theory from Dalian Polytechnic University, Dalian, China, in 2007. He is currently a senior engineer in Shenzhen Smart-Chip Microelectronics Technology Co., Ltd.. His current research interests include power system modeling and power system load forecasting.

Junxiong Ge received the B.S. degrees in enterprise management from Deakin University, Melbourne, Australia, in 2017 and the M.S. in business administration from University of Tasmania, Hobart, Australia, in 2020. He is currently an engineer in Shenzhen Guodian Technology Communication Co., Ltd.. His current research interests include electricity market analysis and power system load forecasting.

Zhanhao Zhang received the B.S. degrees in finance from Deakin University, Melbourne, Australia, in 2017. He is currently an engineer in Laijin (Tianjin) Auto Parts Co., Ltd.. His current research interests include power system modeling and power system load forecasting.

PuKun Lu was born in Changchun City, China. He received the B.S. degree in automation from the Northeast Electric Power University, Jilin, China, in 2020. He is currently pursuing the master’s degree with Northeast Electric Power University, Jilin City. His current research interests include demand response of grid and intelligent power dispatching system.

Shunjiang Wang was born in Yuhuan City, China. He received the Ph.D. degree in computer application technology from the University of Chinese Academy of Sciences, Shenyang, China, in 2019. He is currently a professorate senior engineer in State Grid Liaoning Electric Power Co., Ltd.. His current research interests include distribution system analysis and power system analysis.