www.rericjournal.ait.ac.th

A New Hybrid Model using WTBPNN-EHO and LSTM with EMD-WOA Signal Decomposition for mid and Long-Term Electricity Load Forecasting

Mesbaholdin Salami*, [1], Masoumeh Rostam Niakan*, Masoud Hasani Marzooni*, and Kioumars Heydari*

ARTICLE INFO

Article history:
Received 16 February 2022
Received in revised form
30 August 2022
Accepted 07 November 2022

Keywords:
Deep learning
Electricity demand forecasting
LSTM
Signal decomposition
WOA

ABSTRACT

The manager of the electricity supply chain needs to correctly forecast one of the most important variables affecting the management of the electricity chain. This article has proposed a hybrid model for electricity demand forecasting using deep learning. Firstly, the historical electricity demand data is decomposed using empirical mode decomposition (EMD) algorithm. Then, whale optimization algorithm (WOA) is used to determine signal decomposition levels rounds by EMD and the allocation of signals to neural networks. Parameters that promote accuracy of the forecast are selected using principal component analysis (PCA). A group of signals are fed into a back propagation neural network, whose components are decomposed by wavelet transform. The weights of this neural network are determined by using elephant herding optimization (EHO) algorithm (WTBPNN-EHO). The rest of the signals with higher levels of complexity are fed into the long short-term memory (LSTM) neural network. Finally, the load is calculated by aggregating the results of these two neural networks. Finally, the performance of this model has been compared with other existing models.

1. INTRODUCTION

Electricity load demand is a necessary variable required for planning and controlling the electricity generation and distribution chain [1]. The high accuracy of load forecasting, increases the efficiency of the electricity network planning, and reduces the additional costs caused by excessive production or shortages in the network [2]. In this regard, various models have been designed for forecasting the electricity load demand.

Electricity load demand can be divided into very short, short, midterm and long-term classes in terms of time period [3]. In medium and long-term forecasting, which includes one month and longer durations, economic and social parameters with significant changes in medium and long-term periods are also used in addition to parameters such as temperature and humidity [4]. Also, the load demand forecasting models are structurally divided into three categories: linear, nonlinear, and hybrid models [5].

Linear models are those that employ mathematical formulation to determine the relationship between one or more variables. These include auto regression integrated moving average (ARIMA) [6], autoregressive – moving-average model with exogenous model (ARMAX) [7], linear regression (LR) [8], single exponential (SE) [9], high pass filter model with auto regression (HPF + AR (1)) [10], seasonal auto regression integrated moving average (SARIMA) [11], Fourier seasonal SARIMA (FS-SARIMA) [12] and the like. These models have poor performance with nonlinear signals and are not able to identify complex relationships between parameters and variables [13].

The second category includes nonlinear models that are able to learn, reason, and correct themselves. These models have higher performance with more complex types of signals. Some examples of this category include power load forecasting model using support machine vector and ant colony optimization (SVM-ACO) [14], bees algorithm and artificial neural network (BA-ANN) [15], support vector regression - chaotic artificial with bee colony algorithm (SVM-CABCA) [16], wavelet particle swarm optimization neural network simulation optimization (wavelet-PSO-NNs-SO) [17], chaotic gravitational search algorithm (CGSA) [18], and shared subscribe hyper simulation optimization (SUBHSO) [19]. Getting stuck in the local optimum, and overfitting may occur in these models.

The third category involves hybrid models to cope with the problems of other approaches [20]. In this category, two or more statistical models or, more often, artificial intelligence-based models are used simultaneously. These include empirical mode decomposition minimal redundancy maximal fruit fly optimization general regression neural network (EMD-mRMR-FOA-GRNN) [21], wavelet empirical mode decomposition improved grasshopper optimization algorithm neural network and ARIMA (EMD-IGOA-ARIMA-NN) [22], ARIMA-ANN- multiple linear regression (MLR) (ARIMA-ANN-MLR) [23], and improved EMD ARIMA-neural network fruit fly optimization (ARIMA-NNFOA-IEMD) [24].

*Department of Electricity and Energy Economics, Niroo Research Institute, Tehran, Iran.

[1]Corresponding author:
Email: Mesbah.salami@yahoo.com.

A group of models based on deep learning is included in this category, which have better performance compared to other hybrid models. These include LSTM [25], multi task attention- LSTM (MTAL) [26], neural network- LSTM (NN-LSTM) [27], bidirectional LSTM (BI-LSTM) [28], SARIMA- LSTM [29], convolutional neural network - LSTM (CNN-LSTM) [30], self-attention algorithm-LSTM (SAM-LSTM) [31], and EMD-GA-LSTM [32]. Feature selection methods have also been used to improve performance in electricity load demand forecasting models. These include the seasonal persistence-based regressive model (SPR) [33], two stage mutual information feature selection technique based on the transudative model (MIT-MIT) [34], minimal redundancy maximal relevance (Mrmr) [35], improved multilayer binary firefly algorithm (MBFA) [36], memetic algorithm (MA) [37], and support vector regression (SVR) [38].

This study considered a specific feature selection method in examining the process of developing hybrid models based on deep learning. In order to eliminate the mixing of signals, and increase the accuracy of the forecasting, the input signal is decomposed using EMD. WOA algorithm is used to improve the performance of the EMD by calculating the optimal number of intrinsic mode functions (IMFs) and determining the method of their division and allocation to the two neural networks (EMD-WOA).

Upon determining the optimal number of IMFs and their allocation method, higher- and lower-level IMFs are fed into LSTM and wavelet back propagation neural networks with EHO (WTBPNN-EHO), respectively. Then, the best parameters are selected through PCA and the data are analyzed by two neural networks. LSTM is used to analyze high-level IMFs due to its stronger memory and higher accuracy dealing with more complex signals.

Moreover, the study uses BPNN neural network for increasing the accuracy of forecasts parallel to LSTM to analyze lower-level IMFs. The electricity load consumption signal in BPNN is decomposed using Daubechies of order5 wavelet transform and its weights are optimized through EHO method (WTBPNN-EHO).

As a result, the proposed model of this paper is a combination of EMD, WOA, PCA, WTBPNN-EHO and LSTM models.

The performance of this model is evaluated using a data set of Iran's electricity load consumption. Finally, the accuracy of the proposed model and four deep learning-based models are compared by defining error evaluation criteria.

Furthermore, the validity of the proposed model is compared with the existing models through Wilcoxon method.

The proposed model, and experimental framework are presented in sections 2 and 3, respectively. The conclusions are drawn in the fourth section.

2. PROPOSED MODEL

The proposed model used a deep learning model based on LSTM. In order to promote the performance of LSTM, EMD-WOA method is used for signal decomposition, while WTBPNN-EHO neural network is employed in parallel. The components of the model and its structure are illustrated in the following.

2.1 Model Components

2.1.1 EMD

To obtain the time frequency of a signal, a method should be adopted to extract the latent intrinsic modes of the signal. The EMD algorithm has been target for this purpose. It indicates what frequency exists in a signal at any given moment. It also shows the frequency magnitude. This method is employed for signal decomposition to eliminate signal interference and improve forecasting accuracy [39].

In brief, the EMD algorithm includes the following steps [33].

1- Find the position of all extremum points of signal x(t).
2- Find up and low envelope through a cubic curve.
3- find the average of the up and low curves.
4- Reduce the average of the up and low curves.

$$h_1(x) = y_t - m_1(t) \quad (1)$$

5- Standard deviation (SD) (2) is used as a stop condition:

$$D_k = \frac{\sum_{t=0}^{T} \left| h_1^{k-1}(t) - h_1^{k}(t) \right|^2}{\sum_{t=0}^{T} \left| h_1^{k-1}(t) \right|^2} \quad (2)$$

6- If Equation 2 is not valid, the resultant signal of Step 3 is used as the main signal, and the process then proceeds from Step 1.
7- If Equation 2 is true, the screening process $c_1 = h_1^k(x)$ ends, and is then used as the first IMF, the first upper-level of signal *x(t)*.
8- The remainder is defined as $r_1 = y_t - c_1^k$. When the stop condition is met, it is considered an IMF; otherwise, if it meets Condition *a*, it is considered the primary signal, and Steps 1–4 are repeated to obtain the next IMF. If Condition *a* is not met, it is considered a remainder. Therefore, signal y(t) is defined as Equation 3:

$$y_t = \sum_{n=1}^{N} h_n + r \quad (3)$$

2.1.2 WOA

This algorithm starts with a set of random solutions, in each iteration the search agents update their position according to each of the search agents randomly or with the best solution obtained. This algorithm is performed in three phases, which are as follows:

Encircling prey

The following formulas are used to update search answers and parameters:

$$\vec{H} = \left[\vec{D}.\ \vec{U^*}(t) - \vec{U}(t)\right] \quad (4)$$

$$\vec{U}(t+1) = \vec{U^*}(t) - \vec{B}.\vec{H} \quad (5)$$

In the formula above, D is search space; B and D is coefficients. Also, Y* is the best answer in the t iteration; and $\vec{U}(t)$ is the solutions of the current iteration and $\vec{U}(t+1)$ is the solutions of the next iteration of the algorithm. The vectors B and D are determined as follows:

$$\vec{B} = 2\,\vec{e}.\vec{r} - \vec{a} \quad (6)$$

$$\vec{D} = 2\,\vec{r} \quad (7)$$

Bubble-net attacking method (exploitation phase)

In this mechanism, the distance between the position of the wall and the location of the bait is calculated. Then the spiral movement is created as shown in the following equation:

$$\vec{U}(t+1) = \vec{H'}.e^{bl}.cos(2\pi l) + \vec{U^*}(t) \quad (8)$$

Where H 'is the distance between the bait (the best solution) and the whale. b is a constant. l is a random number in [-1, 1]. We adjust the mathematical model of these two mechanisms. We assume that the probability of 50% choice between these two mechanisms for updating the position of the valves is as follows:

$$\vec{H'} = \left|\vec{U^*}(t) - \vec{U}(t)\right| \quad (9)$$

$$\vec{U}(t+1) = \begin{cases} \vec{U^*}(t) - \vec{B}.\vec{H} & if\ k < 0.5 \\ \vec{H'}.e^{bl}.cos(2\pi l) + \vec{U^*}(t) & if\ k \geq 0.5 \end{cases} \quad (10)$$

Where k is a random number between 0 and 1.

Search for prey (exploration phase)

In the exploration phase, whales in the flock (search agents) randomly search for prey (the best solution) and change their positions according to the positions of other whales. In order to force the search agent away from the reference whale, we use B with values > 1 or < 1. The mathematical model of the exploration stage is as follows:

$$\vec{H} = \left|\vec{D}\,\overrightarrow{Y_{Rand}} - \vec{U}\right| \quad (11)$$

$$\vec{U}(t+1) = \overrightarrow{U_{Rand}} - \vec{B}.\vec{H} \quad (12)$$

In the above equations, U_{Rand} is a randomly-selected position vector (random whale) from the current population

2.1.3. EMD-WOA

In the literature [21]-[22], the number of signal decomposition rounds and the method of their allocation to neural networks are predetermined. In this study, we use WOA meta-heuristic method to optimize these values, to increases the model's accuracy.

Higher-level IMFs are fed into LSTM neural network due its memory and ability of analyzing more complex signals (Figure 1). On the other hand, lower-level IMFs are fed into WTBPNN-EHO neural networks given the high performance of back propagation networks in analyzing signals with less complexity.

For this purpose, the vector $\vec{X}$(t) is calculated using Equation 13.

$$\vec{X}(t) = \left(\theta^t, \theta'^t\right) \quad \forall t \in T \quad (13)$$

Where θ^t is the number of signal decomposition levels in the t iteration, and θ'^t is the signal split point in period t. The fitness function used in this hybrid model is equal to Equation 14.

$$\omega_t = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\partial_{t-1} - \hat{\partial}_{t-1}}{\partial_{t-1}}\right| \quad (14)$$

Table 1. Neural networks parameters.

Neural network parameter name	Quantity
Number of hidden layers n_{hl}	3
The number of neurons in layers (n_{ne}^{i})	(300, 357, 362)
Dimensions of the data set (Batch Size)	100
Algorithm optimizer	adam
Cost Function (Loss)	MSE
Epoch Number (n_{epoch})	150
Activator function (f_{act})	relu
Total number of data (N)	2000
number of Train data (n_{train})	1600
number of Test data (n_{test})	400

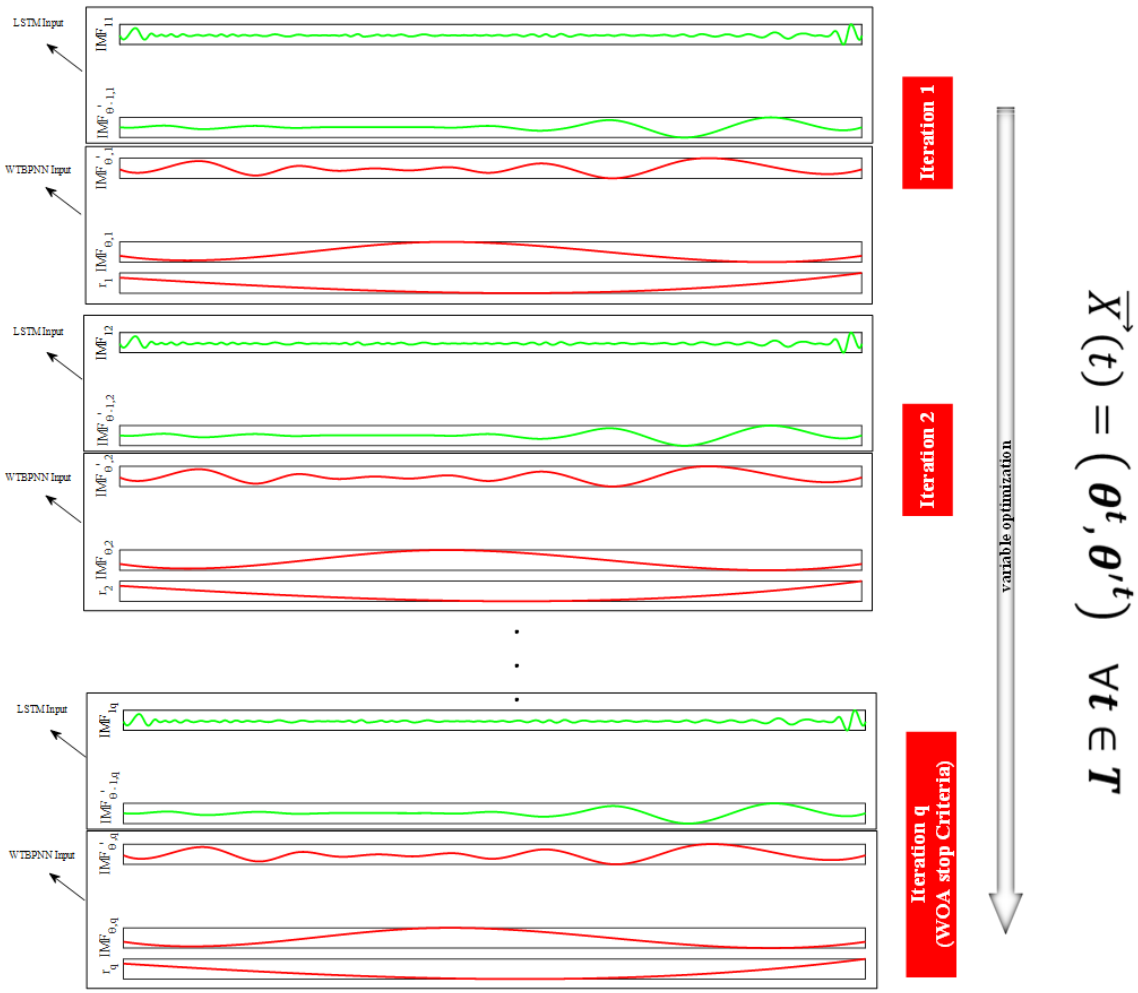


Fig. 1. EMD optimization by WOA.

2.1.4 WTBPNN

BPNN is a deep learning method for a more accurate calculation of the weight gradient, by optimizing a learning algorithm and fixating the weights of neurons through the calculation of the downward gradient of a cost function.

The first step is the feed-forward, multiplying the input data by weights, and then adding the result to the deviation. The output of this step is equal to the actual output.

Then, the loss function indicates the error rate of the feed-forward step, based on which the weights of layers are modified in the second step through the feed-backward [42]. According to many studies of electricity load forecasting, integrating a neural network with wavelet transforms can improve the neural network outputs [43].

One of the spectral analysis approach is to analyze the unstable signal in order to achieve a better time and frequency resolution of the signal. Wavelet transform has two different types, discrete and continuous.

Mathematically, a continuous wavelet transform can be described by the following function:

$$X_w(v,w) = \frac{1}{|v|^2}\int_{\infty}^{-\infty} x(t)L\left(\frac{t-w}{v}\right)dt \qquad (15)$$

In the above formula, $L(t)$ is continuous mother wavelet that has been transferred by factor v and w., there are many mother functions such as Beta, Spline, Shannon, and Hermitian [45-46]. Here, three steps of Daubechies of order 5 (db5) is applied. This study benefits from Mallat *et al.* [44] research that introduced multi regression analysis (MRA). This method consists of two phases: decomposition and a combination of components that is divided into two level of signal frequencies. This process is continued for low frequency signals until meet the termination criterion is met.

Figure 2 displays the decomposition approach represented in the Equation 16.

$$\begin{aligned} D &= l_1+h_1 \\ &= h_1+h_2+l_2 \\ &= h_1+h_2+h_3+l_3 \end{aligned} \qquad (16)$$

Now, we describe three layouts:

- The input layer: The value of the r^{th} node of the first layer is calculated as Y'_r through the multiplication

of the input value Y_i in the vector weight W_{ir} by using F:

$$Y'_r = F\left(\sum_{i=1}^{m} W_{ir}\, Y_i\right) \qquad \forall r = 1.\ldots.p \qquad (17)$$

- The hidden layer: The value of the s^{th} node of the second layer is calculated as Y''_s through the multiplication of the input value by the first-layer node Y'_r by the weight of vector U_{rs} with the help of F':

$$Y''_s = F'\left(\sum_{r=1}^{P} U_{rs}\, Y'_r\right) \qquad \forall s = 1.\ldots.q \qquad (18)$$

- The output layer: The value of the o^{th} node of the output layer is calculated as y_o through the summation of the second-layer input value Y''_s by the weight of vector V_{so} with the help of F''. This layer includes the db5 :

$$Z_o = F''\left(\sum_{s=1}^{q} V_{so}\ Y''_s\right) \qquad \forall o = 1.\ldots.4 \qquad (19)$$

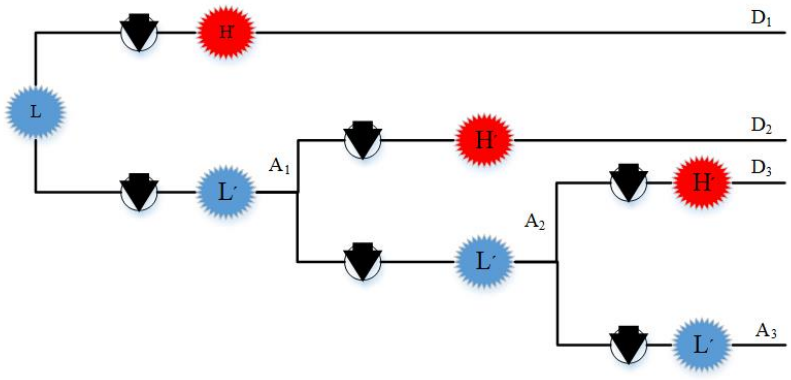


Fig. 2. The three-step decomposition of electricity load demand.

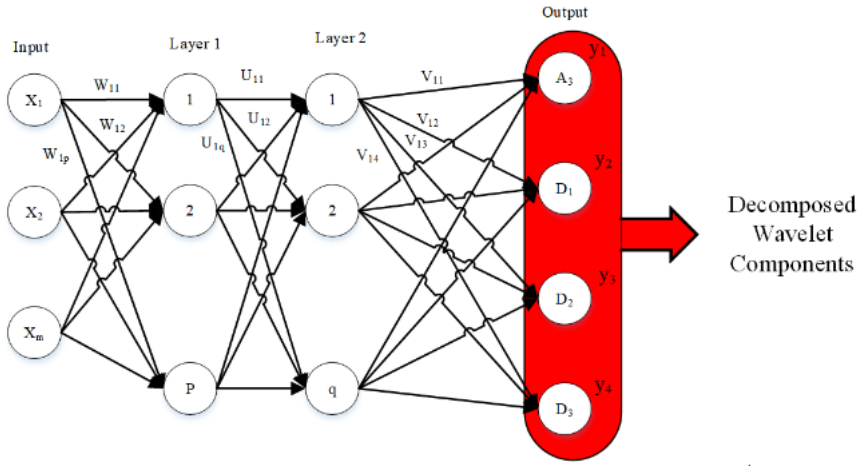


Fig. 3. The neural network structure after propagation through a wavelet transform.

2.1.5. EHO

Elephant herding optimization with two operators has been used as a multifunctional optimization method. A full description of this method is given in [47]. The operators of this algorithm are as follows:

Clan updating operator

In every clan, elephants are under the custodianship of a leader; therefore, an elephant's next position in clan c_i is affected by the leader c_i. The following equation is valid for the jth elephant in the clan c_i:

$$k_{new,c_i,j} = k_{c_i,j} + \alpha \times \left(k_{best,c_i} - k_{c_i,j}\right) \times b \qquad (20)$$

Where $k_{new,c_i,j}$ and $k_{c_i,j}$ refer to the new and previous positions of the j^{th} elephants in the clan c_i, whereas $\alpha = [0,1]$ is a scale coefficient that indicates the effectiveness of the herd leader on the update of other members. Furthermore,b is a random number between 0 and 1, and x_{best,c_i} denotes the clan leader that is the most appropriate response to the optimization problem. The clan leader or the most appropriate elephant of every clan cannot update its position through (21).

$$k_{new,c_i,j} = \beta \times k_{center,c_i} \qquad (21)$$

In (Equation 22) $\beta = [0.1]$ is a scaling coefficient that shows the effectiveness of k_{center,c_i} on $k_{new,c_i,j}$.

According to (21), $k_{new,c_i,j}$ is obtained from all the elephants of the clan c_i through the existing information. It is the center of the clan c_i that can be calculated through the following equation for the next problem:

$$k_{center,c_i} = \frac{1}{n_{c_i}} \times \sum_{j=1}^{n_{c_i}} k_{c_i,j,g} \quad (22)$$

Where $1 \le g \le G$ shows the g^{th} dimension, and n_{c_i} indicates elephent populations in the clan c_i, whereas $k_{c_i,j,d}$ refers to the g^{th} dimension of $k_{c_i,j,g}$.

Segregation operator

In every elephant clan, male elephants leave their families at puberty and start living alone. This separation is modeled as a segregation operator in the EHO algorithm to solve optimization problems. To improve the search power in this algorithm, a solution with the smallest value of the fitness function can be updated through Equation 23 in every generation.

$$k_{worst,c_i} = k_{min} + (k_{max} - k_{min} + 1) \times b \quad (23)$$

In the above equation, k_{max} and k_{min} refer to the maximum and minimum range of elephant's position in the clan c_i, whereas x_{worst,c_i} indicates the worst position that an elephant can have in the tribe. Moreover, b is a random number between 0 and 1.

2.1.6. WTBPNN- EHO

In the previous studies, it has been shown that metaheuristic methods in combination of neural networks have better results rather than other exact solution methods.[3] Many different metaheuristic algorithms such as ACO, Differential Evolution (DE) , GA and etc. have been used to improve neural network performance.

This study uses advanced EHO algorithm for this purpose due to its higher accuracy in achieving optimal solutions compared to other metaheuristic algorithms. [47] Optimization of neural network weights were conducted through the following steps:

- Adjusting the algorithm's parameters such as the number of layout, nodes, size of EHO population and etc.
- Fitness function is considered equal to the correlation coefficient. (Equation 24).

$$minimize \left(\frac{Covariance(x,y)}{\sqrt{variance(x) vararriance(y)}} \right) \quad (24)$$

- For creation the initial coordinates for the population, the motion interval of x_i^t is defined as (25), which is discrete; therefore, the initial coordinates are defined as (26).

$$D_{x_{c_i,j}} = \left[\min(x_{c_i,j}) . \max (x_{c_i,j}) \right] \quad (25)$$

$$x_{c_i,j}{}^0 = Rand \left(D_{x_{c_i,j}} \right) \quad (26)$$

-The motion function of the population in every iteration should be defined as Equations 25 and 26. The new solutions are generated until the termination condition is met.

The forecasted db5 components that generated by the neural network sum up to calculate$\widehat{S}_t$. Figure 4 shows the combination of signals to calculate the final forecasted electricity load demand.

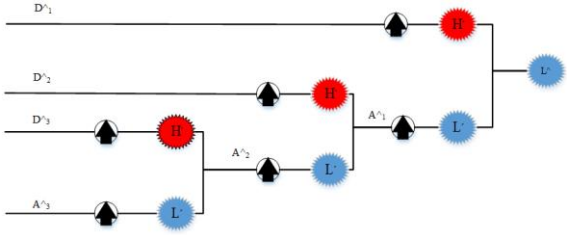


Fig 4. combination of the electricity load demand component.

The following fitness function should be minimized in the neural network:

$$f_t = \frac{1}{N} \sum_{t=1}^{N} (S_t - \widehat{S}_t)^2 \quad (27)$$

Where S_t refers to the real and $\widehat{S}_t$ denotes the forecasted electricity load at t.

2.1.7. LSTM

LSTM networks use a memory cell with feedback gain to eliminate successive matrix multiplication. [48] The model is shown in Figure 5. LSTM network consists of different components:

Input Gate:
This gate recognizes which of the values is suitable for memory improvement and assigns weight to the stored values-based on the tan function between 0 to 1.

$$i_t = \sigma (E_i . [S_{t-1}, Z_t] + q_i) \quad (28)$$

$$\tilde{C}_t = tanh (E_c . [S_{t-1}, Z_t] + q_c) \quad (29)$$

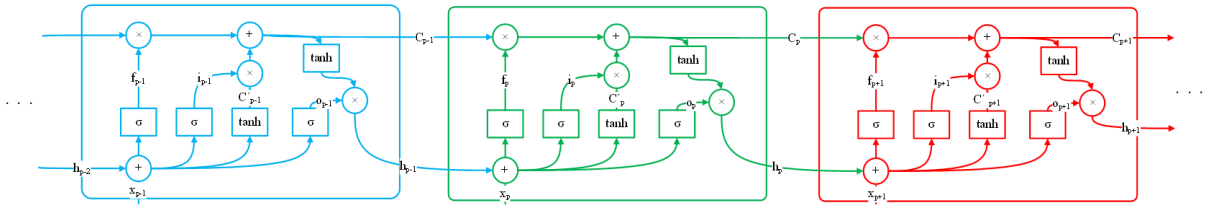


Fig. 5. LSTM model diagram.

Forgetting Gate:
In this gate, the sigmoid function keeps or deletes the previous input information (h_{t-1})) and checks the content input (X_t) in the C_{t-1}cell state :

$$f_t = \sigma\left(E_f.[S_{t-1}, Z_t] + q_f \right) \quad (30)$$

Output Gate:
The Sigmoid function retains or deletes the values-based on the input data and the previous memory using the tan function and assigns weight to the retained values.

$$u_t = \sigma\left(E_0.[S_{t-1}, Z_t] + q_0 \right) \quad (31)$$

$$S_t = u_t.\tanh(C_t) \quad (32)$$

2.2. Model structure

The proposed model uses two parallel WTBPNN neural networks optimized through EHO and LSTM methods to forecast electricity load demand. The data is analyzed through EMD before being fed to the neural networks. The WOA method is used to determine signal decomposition levels and the method of their allocation to neural networks. The input parameters of the neural network are then selected by PCA method. Finally, the results of neural network analysis are aggregated. Figure 6 shows the structure of the proposed model. The model is implemented taking five steps as follows:

1. The input signal is decomposed by emd. The WOA algorithm determines the number of signal decomposition rounds and the method of their allocation to neural networks.
2. The best parameters are selected though PCA.
3. The signals in which $IMF_{pq} > IMF_{\theta q}\ \forall q \in$ Q are fed into WTBPNN-EHO
4. The signals in which $IMF_{pq} \leq IMF_{\theta q}\ \forall q \in Q$ are fed into LSTM
5. LSTM and WTBPNN- EHO forecasted value are sum up to achieve final forecasting value

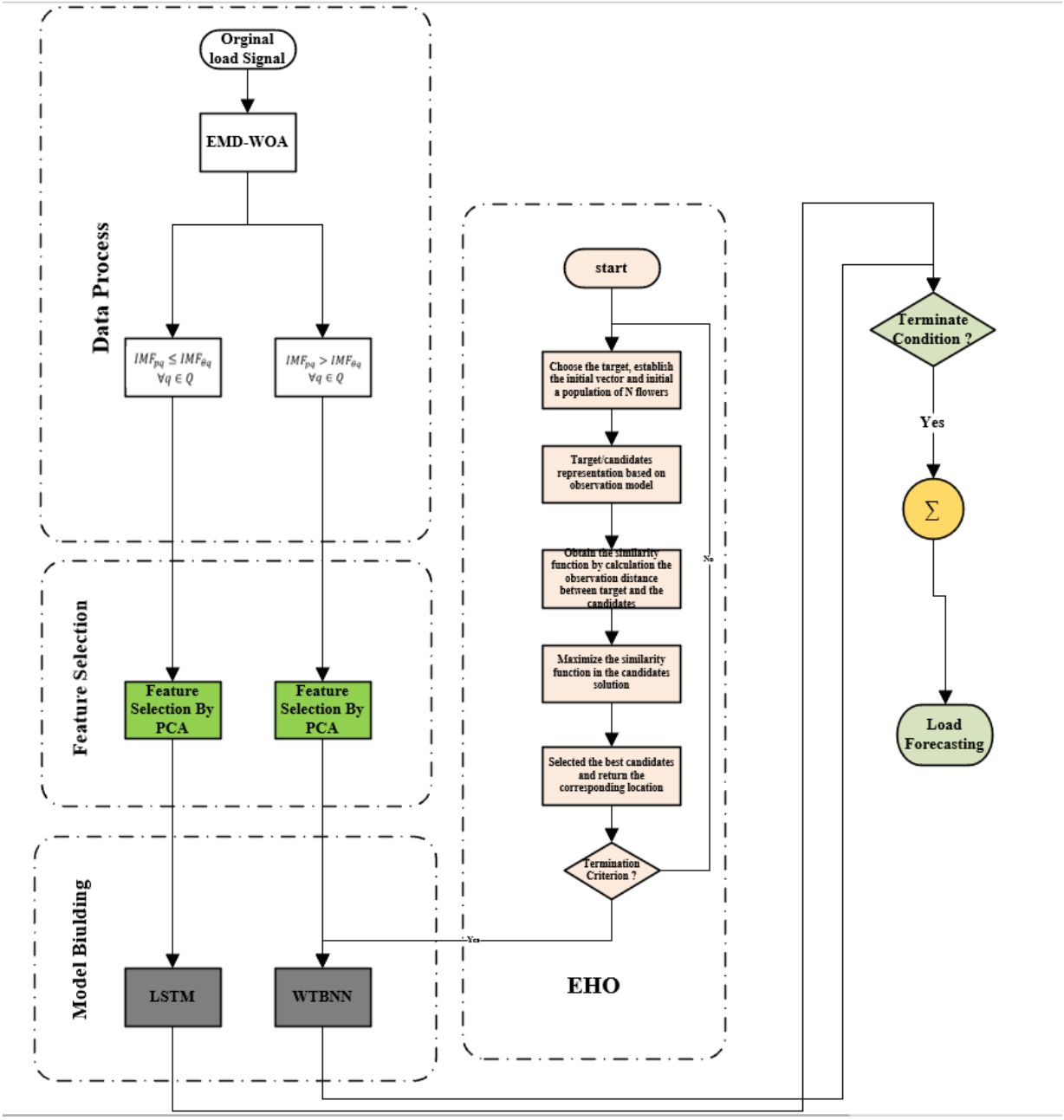


Fig. 6. proposed model structure.

3. EXPERIMENTAL FRAMEWORK

3.1 Error Criteria

The mean absolute percentage error (MAPE) in (33), the mean absolute error (MAE) in (34), and the root mean square error (RMSE) in (35) are employed to compare the performance of the proposed model with the other models, and assess forecasting improvements made by the methods introduced in [49]. These equations are used in Section 4 to evaluate performance of proposed model.

$$MAPE = \frac{1}{M}\sum_{t=1}^{M}\left|\frac{\vartheta_t - \widehat{\vartheta_t}}{\vartheta_t}\right| \quad (33)$$

$$MAE = \frac{1}{M}\sum_{t=1}^{M}\left|\vartheta_t - \widehat{\vartheta_t}\right| \quad (34)$$

$$RMSE = \frac{1}{M}\sqrt{\sum_{t=1}^{M}\left(\vartheta_t - \widehat{\vartheta_t}\right)^2} \quad (35)$$

3.2 Discussion and Results

This model is compared with deep-learning-based hybrid models, including SARIMA-LSTM, NN-LSTM, SAM-LSTM and CNN-LSTM in terms of accuracy. The model uses best parameters among population, air temperature, and historical data on electricity load demand in previous cycles, humidity, and power consumption peaks per day. The signal of the electricity load demand historical data is shown in Figure 7.

For instance, we used this model to forecast January 10, 2016 demand. Figure 8 indicates the signal decomposition of historical data of electricity load demand through the EMD method in different iterations of the WOA.

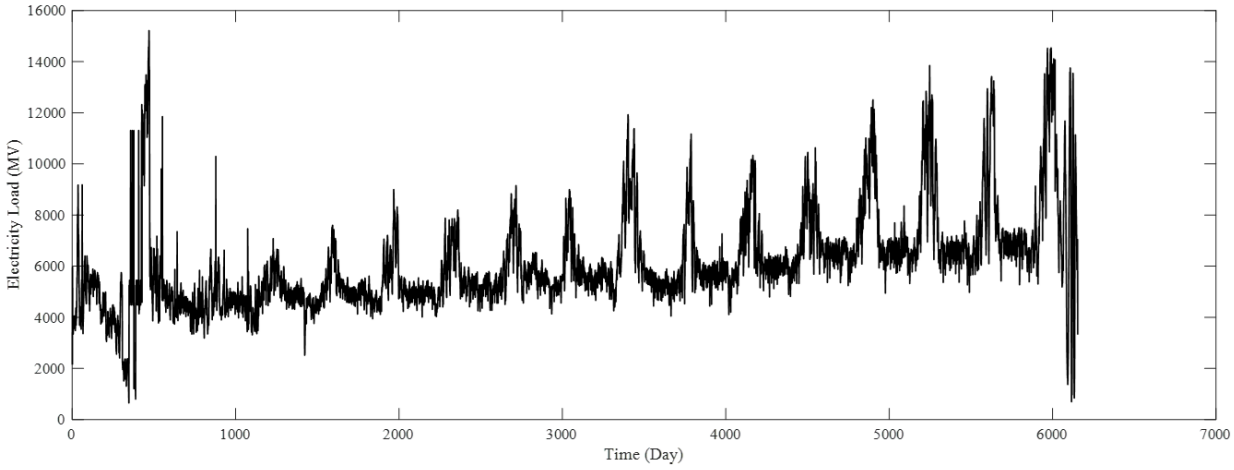


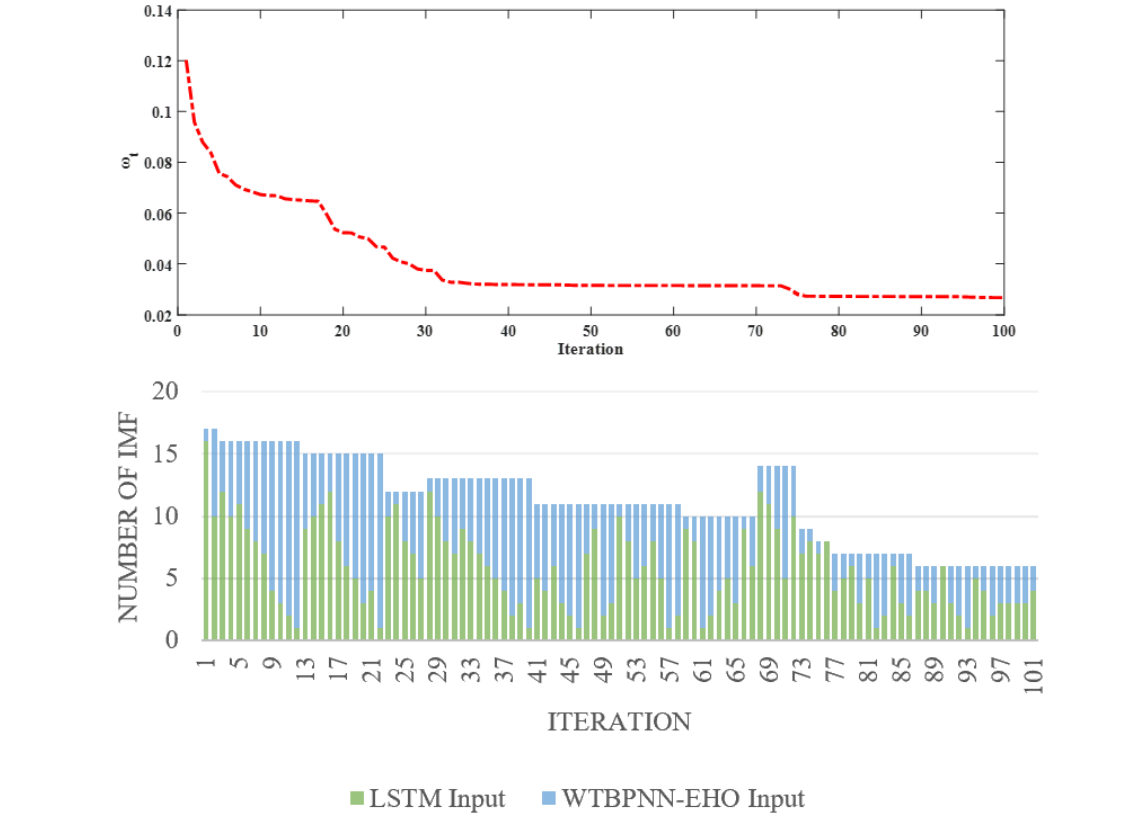
Fig. 7. Historical data of electricity load consumption.

The selection of features is also done by the PCA model, which shows how the model works in Figure 9.

Steps 3 and 4 of Section 2.2 that forecasting the amount of electricity load are shown in Figure 10.

The results of load forecasting on January 10, 2016 by the proposed model and the four existing models are shown in Figure 12, part a. In this diagram, the actual amount of load demand on this day is equal to 3647 and the values forecasting by SARIMA-LSTM, NN-LSTM, SAM-LSTM and CNN-LSTM algorithms and the proposed model are equal to 4011, 3974, 3829, 3574 and 3610 MW respectively. Also shown in Figure 11 are the other days of 2016 by this model.

Table 2 is a comparison of the proposed model with the existing models based on the error assessment criteria. To compare the performance of the models, the standard error chart for real and forecasts values is shown in Figure 1. Table 2 indicates the results of executing the proposed model with those of the other models based on the MAPE on average for 2016. Accordingly, the proposed model improved accuracy by 6.9% compared to the SARIMA-LSTM; however, it improved accuracy by 4.9% instead of the nonlinear NN-LSTM. Compared with the SAM-LSTM, the proposed model improved accuracy by 1.4%, whereas it improved accuracy by 0.3% compared to CNN-LSTM. The proposed model was also employed to run similar analyses for the 2017–2019 period, which indicated that the model managed to maintain the improvements for the data of the other three years. Figure 13 demonstrates the MAPE values for the four study years.



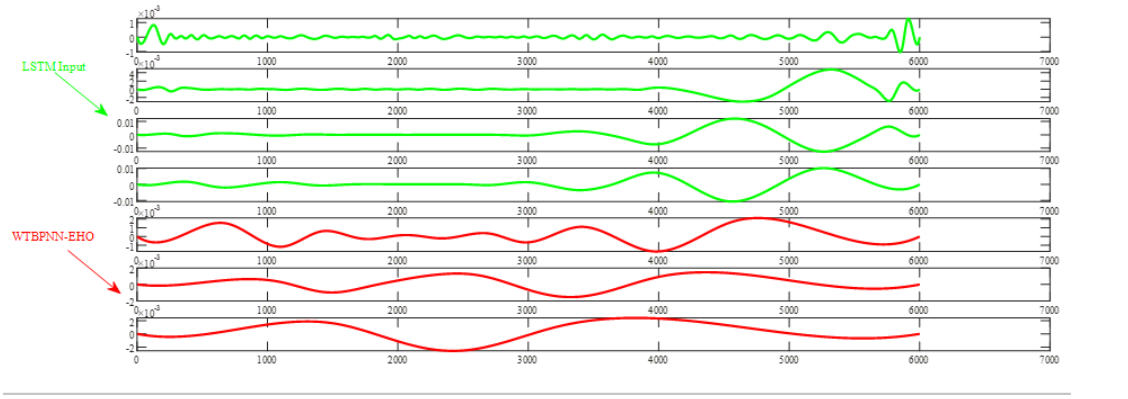


Fig. 8. Signal decomposition with EMD-WOA.

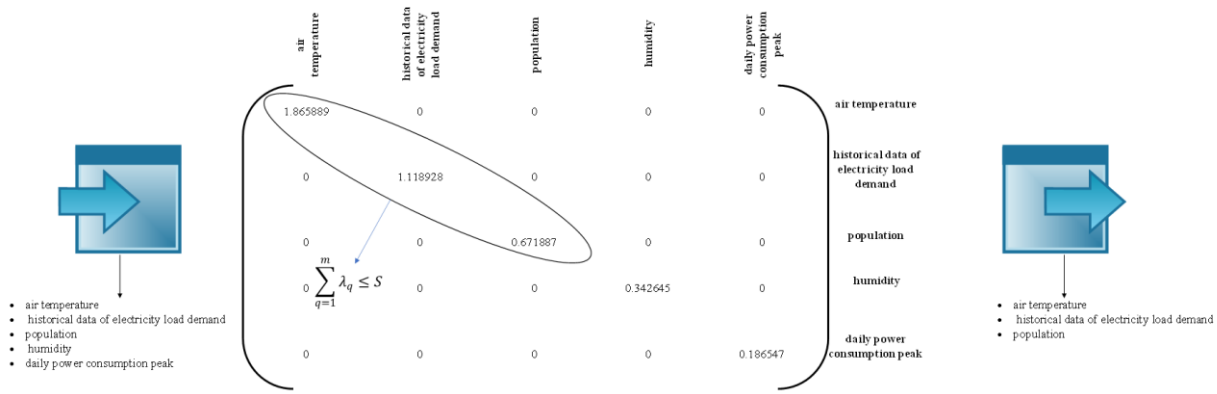


Fig. 9. Feature selection by PCA model.

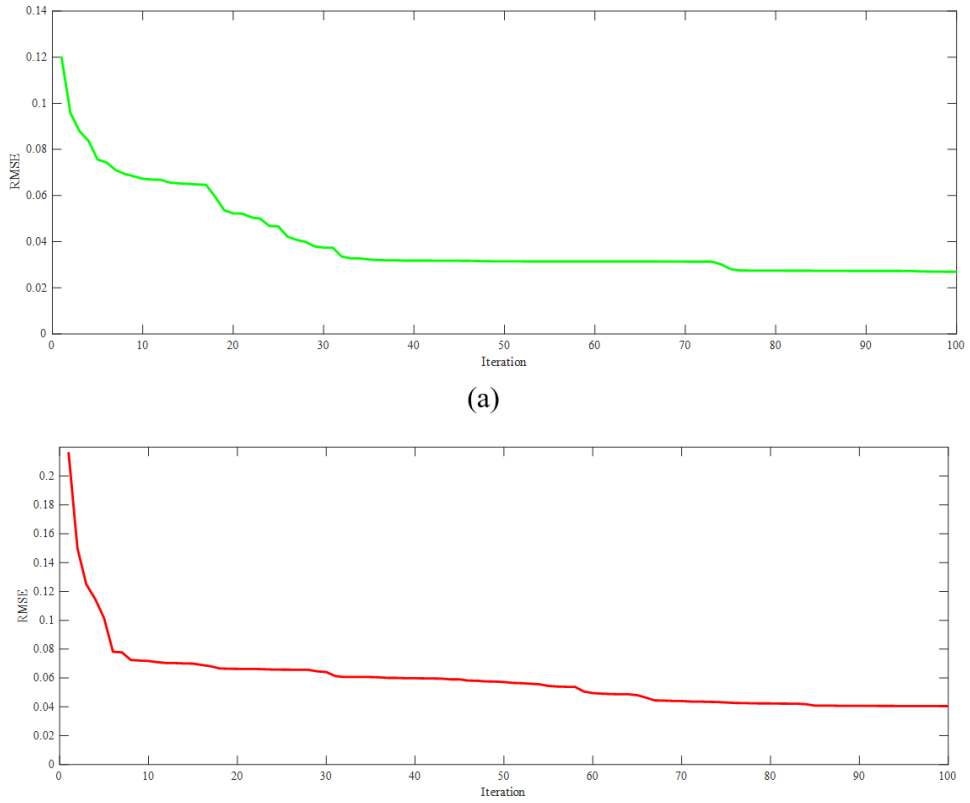


Fig. 10. RMSE optimization during execution a) WTBPNN-EHO b) LSTM.

Table 2. Comparing the performance of the proposed model with those of other models in error assessment index.

Model	Criterion	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	Average
SARIMA-LSTM	MAPE	0.099	0.096	0.096	0.107	0.100	0.098	0.092	0.102	0.096	0.101	0.096	0.099	0.098
	MAE (MW)	326.71	346.57	329.45	335.88	368.81	347.83	326.88	307.63	354.21	327.16	353.98	325.72	337.57
	RMSE	40.12	43.61	41.31	41.98	44.49	44.74	40.05	38.23	43.38	40.39	43.88	40.51	41.89
NN-LSTM	MAPE (%)	0.074	0.079	0.075	0.082	0.087	0.086	0.072	0.070	0.078	0.080	0.078	0.074	0.078
	MAE (MW)	284.46	284.43	303.93	286.30	316.48	339.61	333.27	270.79	266.57	295.62	300.93	295.07	298.12
	RMSE	36.24	35.95	37.47	35.68	38.80	40.73	40.83	33.09	33.87	35.94	35.90	35.79	36.69
SAM-LSTM	MAPE (%)	0.042	0.045	0.043	0.042	0.044	0.041	0.044	0.045	0.045	0.037	0.042	0.042	0.043
	MAE (MW)	118.69	133.17	124.61	121.54	129.83	117.82	126.26	132.61	133.83	103.10	125.95	122.63	124.17
	RMSE	14.68	16.12	15.42	15.11	15.51	14.79	15.45	16.24	16.32	13.13	15.62	15.23	15.30
CNN-LSTM	MAPE (%)	0.029	0.032	0.034	0.032	0.031	0.032	0.033	0.031	0.032	0.031	0.029	0.032	0.032
	MAE (MW)	88.05	73.31	82.92	91.42	81.70	78.98	85.96	88.26	81.43	86.11	82.33	82.48	83.58
	RMSE	10.65	9.36	10.42	11.42	10.22	9.99	10.87	11.05	10.42	11.03	10.62	10.36	10.53
proposed model	MAPE (%)	0.0298	0.0304	0.0289	0.0284	0.0286	0.0298	0.0284	0.0288	0.0271	0.0288	0.0288	0.0301	0.029
	MAE (MW)	48.85	51.12	44.65	44.29	44.20	48.85	43.65	45.41	37.73	45.63	45.64	49.50	45.79
	RMSE	6.00	6.34	5.74	5.76	5.78	6.01	5.60	5.81	5.20	5.76	5.86	6.13	5.83

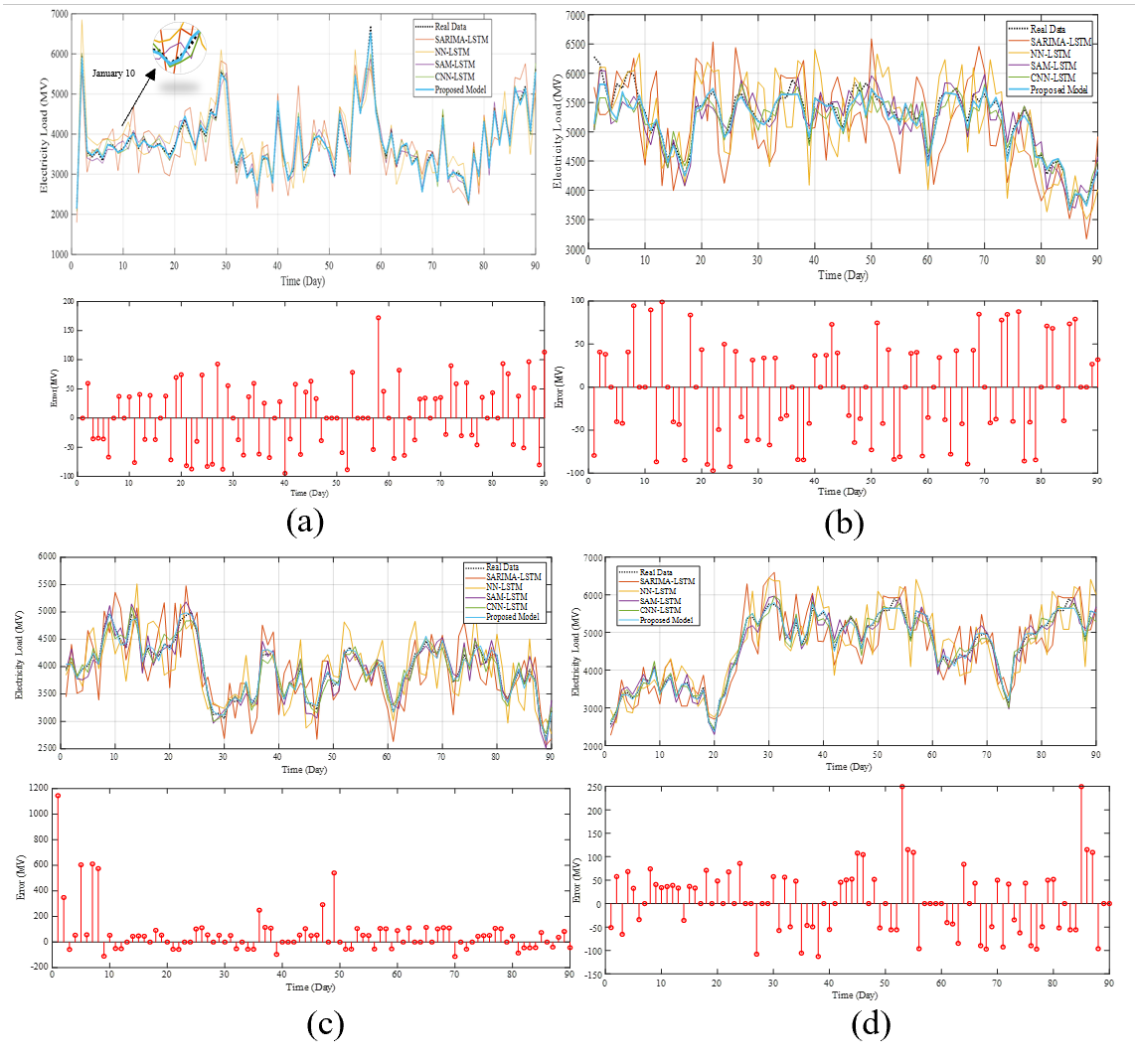


Fig. 11. Comparison of performance of models in a) spring b) summer c) autumn d) winter 2016.

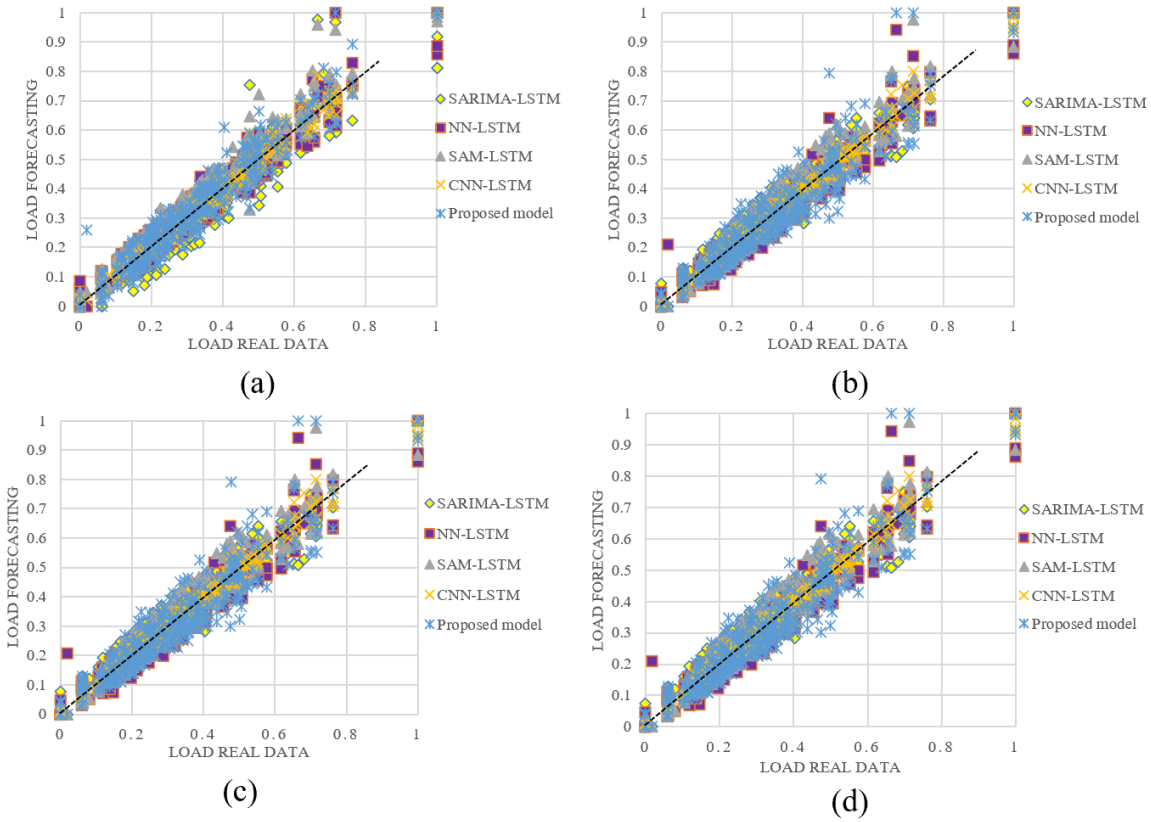


Figure 12. Standard error value of the proposed model with other models in a) spring b) summer c) autumn d) winter 2016.

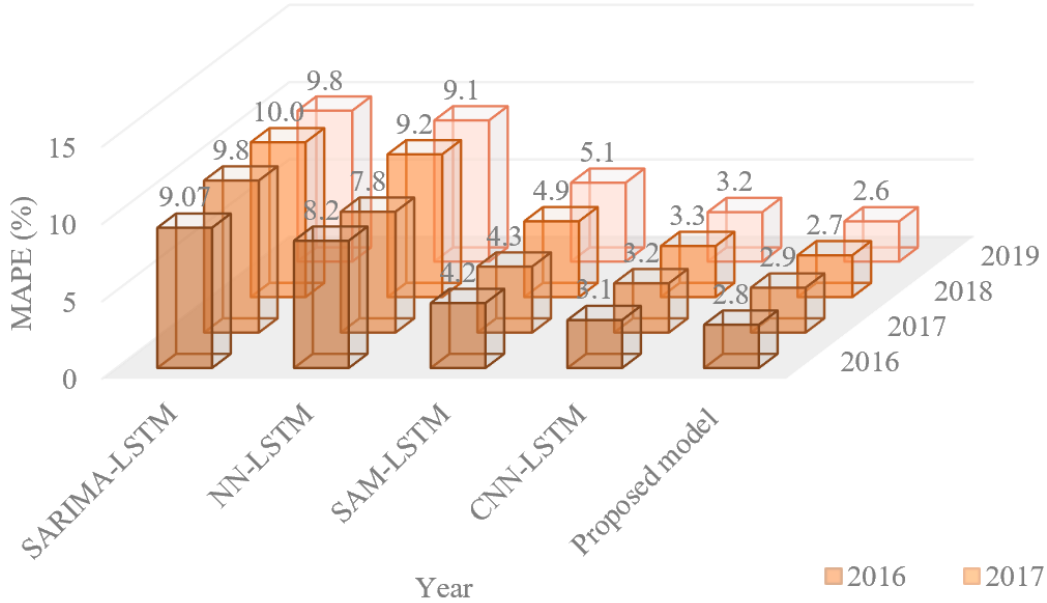


Fig. 13. Comparison models in MAPE for the 2016–2019 period.

3.3 Validation

If random pair values are considered with the sample size of n, the number of values will be 2n due to the odd number of samples. Based on the components, the pairwise observations are shown as $y_{1,i}$ and $y_{2,i}$ for $i = 1\ to\ n$. Statistical hypothesis testing aims to determine the differences between two pairs. Therefore, the null hypothesis is considered the equality of values between pairs, whereas the alternative hypothesis considers a difference between pairs [50]. Accordingly, the following steps are taken for hypothesis testing through the Wilcoxon method:

I. The absolute differences of pairs are calculated as $|x_{2,i} - x_{1,i}|$ for $i = 1,2,\dots,n$ for all observations, and the signs (positivity and negativity) of these differences were determined through $sgn(x_{2,i} - x_{1,i})$.

II. The pairs are excluded from the analysis when their sign functions were zero. Therefore, n_r remaining pairs are considered. However, such pairs can be used in only positive or negative groups to prevent a substantial reduction in the number of observations when the sample is too small.
III. The resultant values of absolute differences ($|x_{2,i} - x_{1,i}|$) are sorted from low to high.
IV. The ranks of absolute values of differences are determined. Some knotted ranks might have been obtained from some values. In this case, the mean ranks are considered for such observations. The smallest difference is ranked first, and the different rank of pair *i* is shown by R_i.
V. The *F* statistic is calculated through Equation 36.

$$\mathrm{F} = \sum_{i=1}^{n_r} [sgn\left(x_{2,i} - x_{1,i}\right) . R_i] \quad (36)$$

That F is known as the total signed ranks. Under the null hypothesis, the probability distribution *W* is complicated; however, its mean equals zero, and its variance is obtained from the following equation:

$$\sigma_w^2 = \frac{n_r\,(n_r + 1)(2n_r + 1)}{24} \quad (37)$$

Despite standard tables in this method, the null hypothesis is rejected if $|\mathrm{F}| < \mathrm{F}_{\alpha,nr}$. Table 3 shows the validity of results between proposed model with others.

Table 3. Wilcoxon signed-rank test.

Compared Model	Wilcoxon signed-rank test	
	α=0.025, F=2	α=0.05, F=3
SARIMA	0	0
BAANN	2	2
IEMDFOA	3	3
proposed model without WOA	2	2

4. CONCLUSIONS AND SUGGESTIONS

This paper follows the development of medium- and long-term forecast models to determine load forecasting that is helpful for management of the electricity supply chain. We used a deep-learning-based hybrid model to find a best result for demand variable compared with last models.

Historical data on electricity load consumption have been broken down into sub-signals using EMD. In this model, a smart mechanism has been employed to use WOA for determining the optimum number of signal decomposition rounds by EMD, as well as the method of feeding decomposed signals to each of the two neural networks. The best parameters contributing to the load demand have been selected from the parameters available in the database through PCA.
WTBPNN-EHO method has been used to increase the accuracy of electricity load demand forecasting, whose input components have been decomposed by wavelet transform method and its weights were optimized by EHO.

Also, LSTM has been used in parallel. The results of each of the two neural networks have been aggregated to obtain the final forecasted value. The performance comparison of this model with other existing models based on the data of Iran network indicated an improvement in the accuracy of electricity load forecasting in the proposed model.

We recommend development of a mechanism for selecting the factors in addition to features in future research. It is also recommended that a smart method be considered to optimize the parameters of neural networks simultaneously.

REFERENCES

[1] Salami M, Movahedi Sobhani. F., and Ghazizadeh. M.S., 2020. Evaluating power consumption model and load deficit at different temperatures using clustering techniques and presenting a strategy for changing production management. *IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, pp. 341-345, doi: 10.1109/DSMP47368.2020.9204247.

[2] Salami M., Movahedi Sobhani F., and Ghazizadeh M.S., 2018. Multi-objective model for fair pricing of electricity using the parameters from the iran electricity market big data analysis. *International Journal of Industrial Mathematics* 10(4): 359-374.

[3] Debnath K.B. and M. Mourshed. 2018. Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews* 88: 297-325, https://doi.org/10.1016/j.rser.2018.02.002.

[4] Chen G., Tang B., Zeng X., Zhou P., Kang P., and Long H., 2022. Short-term wind speed forecasting based on long short-term memory and improved BP neural network. *International Journal of Electrical Power & Energy Systems* 134: 107365. https://doi.org/10.1016/j.ijepes.2021.107365.

[5] Zhang Z., Wu X., Liao S., and Cheng C., 2022. An ultra-short-term scheduling model for cascade hydropower regulated by multilevel dispatch centers suppressing wind power volatility. *International Journal of Electrical Power & Energy Systems* 134: 107467. https://doi.org/10.1016/j.ijepes.2021.107467.

[6] Dudek G., 2016. Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research* 130: 139-14.

[7] Pao H.T., 2006. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. *Energy* 31(12): 2129-2141. https://doi.org/10.1016/j.energy.2005.08.010.

[8] Lee C.M. and C.N. Ko. 2011. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications* 38(5): 5902-5911.

[9] Cadenas E., Jaramillo O.A., and Rivera W., 2010. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single

exponential smoothing method. *Renewable Energy* 35(5): 925-930.

[10] Saab S., Badr E., and Nasr G., 2001. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. *Energy* 26(1): 1-14. https://doi.org/10.1016/S0360-5442(00)00049-9.

[11] Zhu S., Wang J., Zhao W., and Wang J., 2011. A seasonal hybrid procedure for electricity demand forecasting in China. *Applied Energy* (88)11: 3807-3815.https://doi.org/10.1016/j.apenergy.2011.05.005.

[12] Wang Y., Wang J., Zhao G., and Dong Y., 2012. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy* 48: 284-294.

[13] Hickey E., Loomis D.G., and Mohammadi H., 2011. Forecasting hourly electricity prices using ARMAX–GARCH models: An application to MISO hubs. *Energy Economics* 34(1): 307-315. https://doi.org/10.1016/j.eneco.2011.11.011.

[14] Niu D., Wang Y., and Dash Wu D., 2010. Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications* (37)3: 2531-2539. https://doi.org/10.1016/j.eswa.2009.08.019.

[15] Behrang M.A., Assareh E., Assari M.R., and Ghanbarzadeh A., 2011. Using bees algorithm and artificial neural network to forecast world carbon dioxide emission. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects* 33(19): 1747-1759.

[16] Hong W.C., 2011. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* 36(9): 5568-5578. https://doi.org/10.1016/j.energy.2011.07.015.

[17] Salami M.; Movahedi Sobhani F.; and Ghazizadeh, M.S., 2018. Short-term forecasting of electricity supply and demand by using the wavelet-PSO-NNs-SO technique for searching in big data of Iran's electricity market. *Data* 3, 43. https://doi.org/10.3390/data3040043

[18] Ju F.Y. and W.C. Hong. 2013. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. *Applied Mathematical Modelling* 37(23): 9643-9651. https://doi.org/10.1016/j.apm.2013.05.016.

[19] Salami M., Movahedi Sobhani F., Ghazizadeh M.S., and Sadegh M., 2019. Shared subscribe hyper simulation optimization (SUBHSO) algorithm for clustering big data – using big databases of Iran electricity market. *Applied Computer Systems* 24(1): 49-60. https://doi.org/10.2478/acss-2019-0007.

[20] Guo X., Gao Y., Li Y., Zheng D., and Shan D., 2021. Short-term household load forecasting based on Long- and Short-term time-series network, *Energy Reports* 7(1): 58-64. https://doi.org/10.1016/j.egyr.2021.02.023.

[21] Liang Y., Niu D., and Hong W.C., 2019. Short term load forecasting based on feature extraction and improved general regression neural network model. *Energy* 166: 653-663. https://doi.org/10.1016/j.energy.2018.10.119.

[22] Salami M., Movahedi Sobhani F., and Ghazizadeh M.S., 2020. A hybrid short-term load forecasting model developed by factor and feature selection algorithms using improved grasshopper optimization algorithm and principal component analysis. *Electrical Engineering* 102: 437–460. https://doi.org/10.1007/s00202-019-00886-7.

[23] Díaz-Robles L.A., Ortega J.C., Fu J.S., Reed G.D., Chow J.C., Watson J.G., and Moncada-Herrera J.A., 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment* 42(35): 8331-8340, https://doi.org/10.1016/j.atmosenv.2008.07.020.

[24] Zhang J., Wei Y.M., Li D., Tan Z., and Zhou J., 2018. Short term electricity load forecasting using a hybrid model. *Energy* 158: 774-781. https://doi.org/10.1016/j.energy.2018.06.012.

[25] Wang Y., Gan D., Sun M., Zhang N., Lu Z., and Kang C., 2019. Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy* 235: 10-20. https://doi.org/10.1016/j.apenergy.2018.10.078.

[26] Qin J., Zhang Y., Fan S., Hu X., Huang Y., Lu Z., and Liu Y., 2022. Multi-task short-term reactive and active load forecasting method based on attention-LSTM model. *International Journal of Electrical Power & Energy Systems* 135: 107517, https://doi.org/10.1016/j.ijepes.2021.107517.

[27] Memarzadeh G. and F. Keynia. 2021. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. *Electric Power Systems Research* 192: 106995, https://doi.org/10.1016/j.epsr.2020.106995.

[28] Mughees N., Mohsin S.Y., Mughees A., and Mughees A., 2021. Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting. *Expert Systems with Applications* 175: 114844.

[29] Kumar Dubey A., Kumar A., García-Díaz V., Sharma A. K., and Kanhaiya K., 2021. Study and analysis of SARIMA and LSTM in forecasting time series data. *Sustainable Energy Technologies and Assessments* 47: 101474. https://doi.org/10.1016/j.seta.2021.101474.

[30] Guo X., Zhao Q., Zheng D., Ning Y., Gao Y., 2020. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. *Energy Reports* 6(9): 1046-1053. https://doi.org/10.1016/j.egyr.2020.11.078.

[31] Zang H., Xu R., Cheng L., Ding T., Liu L., Wei Z., Sun G., 2021. Residential load forecasting based on LSTM fusing self-attention mechanism with pooling. *Energy* 229: 120682. https://doi.org/10.1016/j.energy.2021.120682.

[32] Salami M., Niakan M.R. and Hasani M., 2021. Electricity load forecasting: a new hybrid model

combining wavelet transform, genetic algorithm and LSTM models. In *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pp. 1-6, doi: 10.1109/ICECCME52200.2021.9591091.

[33] Kychkin A.V. and G.C. Chasparis. 2021. Feature and model selection for day-ahead electricity-load forecasting in residential buildings. *Energy and Buildings* 249: 111200.

[34] Ghadimi N., Akbarimajd A., Shayeghi H., Abedinia O., 2018. Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting. *Energy* 161: 130-142.

[35] Dai Y. and P. Zhao. 2020. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Applied Energy* 279: 115332.

[36] Xie W., Wang L., Yu K., Shi T., and Li W., 2023. Improved multi-layer binary firefly algorithm for optimizing feature selection and classification of microarray data. *Biomedical Signal Processing and Control* 79(Part 1): 104080.

[37] Hu Z., Bao Y., Chiong R., and Xiong T., 2015. Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. *Energy* 84: 419-431. https://doi.org/10.1016/j.energy.2015.03.054.

[38] Hu Z., Bao Y., Chiong R., and Xiong T., 2015. Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection. *Energy* 84: 419-431. https://doi.org/10.1016/j.energy.2015.03.054.

[39] Sungkono, Bahri A.S., Warnana D.D., Monteiro Santos F.A., Santosa B.J., 2014. Fast, simultaneous and robust VLF-EM data denoising and reconstruction via multivariate empirical mode decomposition. *Computers & Geosciences* 67: 125-138. https://doi.org/10.1016/j.cageo.2014.03.007.

[40] Tawhid M.A. and A.M. Ibrahim. 2021. Solving nonlinear systems and unconstrained optimization problems by hybridizing whale optimization algorithm and flower pollination algorithm, *Mathematics and Computers in Simulation* 190: 1342-1369.

[41] Lee L.C. and A.A. Jemain. 2021. On overview of PCA application strategy in processing high dimensionality forensic data. *Microchemical Journal* 169: 106608.

[42] Chen J., Rong Y., Zhu Q., Chandra B., and Zhong H., 2021. A generalized minimal residual based iterative back propagation algorithm for polynomial nonlinear models. *Systems & Control Letters* 153: 104966.

[43] Mallat S., 1989. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 11(7): 674-693.

[44] Conejo A., Plazas M., Espinola R., Molina A., 2005. Day-ahead electricity price forecasting using wavelet transform and ARIMA models. *IEEE Transaction on Power Systems* 20(2): 1035-1042.

[45] Reis A. and A. da Silva. 2005. Feature extraction via multiresolution analysis for short term load forecasting. *IEEE Transaction on Power Systems* 20(1): 189-198.

[46] Bowden N. and J.E. Payne. 2008. Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics* 30(6) 3186-3197.

[47] Tawhid M.A. and A.M. Ibrahim. 2021. Solving nonlinear systems and unconstrained optimization problems by hybridizing whale optimization algorithm and flower pollination algorithm. *Mathematics and Computers in Simulation* 190: 1342-1369.

[48] Nourani V. and N. Behfar. 2021. Multi-station runoff-sediment modeling using seasonal LSTM models. *Journal of Hydrology* 601: 126672.

[49] Willmott C.J., 1982. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63: 1309–1313.

[50] Bagkavos D. and P.N. Patil. 2021. Improving the Wilcoxon signed rank test by a kernel smooth probability integral transformation. *Statistics & Probability Letters* 171: 109026. https://doi.org/10.1016/j.spl.2020.109026.