Electricity price and load demand forecasting using an adaptive hybrid BiLSTM model based on wavelet transform technique and Pareto optimization: an application in the smart cities

MOHAMMAD ASADPOUR[1,*], **NAZILA POURHAJI**[1], **AND ALI AHMADIAN** [2]

[1] *Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran*
[2] *Department of Electrical Engineering, University of Bonab, Bonab, Iran*
[*] *Corresponding author email: m_asadpour@tabrizu.ac.ir*

Manuscript received 02 September, 2023; revised 01 October, 2023; accepted 24 October, 2023. Paper no. JEMT-2309-1467.

Accurate forecasting of electricity price and load demand is an essential requirement for managing energy production and consumption in a smart city. In this paper, an adaptive hybrid model is presented for accurate short-term forecasting of electricity price and load demand based on wavelet transform (WT) decomposition, mutual information (MI) and interaction gain (IG) feature selection methods, and Pareto optimization technique with BiLSTM network called WT-MI-IG-BiLSTM. In this model, first, the electricity price and load demand signals are decomposed using the WT technique. Then, the variables that have the most excellent effect on the prediction are selected by the MI and IG feature selection method. In the forecasting stage, prediction is made with the BiLSTM network, and the combination of networks prediction vectors provides the final prediction result. PJM electricity market price and load demand data sets in 2006 and 2018 and five error metrics including RMSE, MAE, MAPE, Variance, and R-Squared are used to evaluate the model. To demonstrate the high capability of the WT-MI-IG-BiLSTM model, the proposed model has been compared with the MI-IG-BiLSTM, WT-MI-IG-LSTM, and MI-IG-LSTM models. Based on the obtained results, the proposed WT-MI-IG-BiLSTM model compared to the MI-IG-BiLSTM model, which is the best benchmark model, has 17-18.16% improvement in accuracy of electricity price forecasting and 21.8% in accuracy of electricity load forecasting. Finally, the Pareto optimization algorithm has implemented on the model, and a set of optimal models with optimal accuracy and execution time has presented in the Pareto front chart.

keywords: *Wavelet transform, Mutual information, Interaction gain, LSTM, BiLSTM.*

http://dx.doi.org/10.22109/JEMT.2023.414599.1467

NOMENCLATURE

Parameters

d Number of solutions set in optimization
m_f Number of functions in optimization
m_t Number of variables
n Total number of input data
n_h Total number of hidden units
n_o Total number of outputs
nt Total number of test data
n_t Number of samples
r Dimension of each input data sequence
SS The standard deviation of sample data
$\bar{X}$ The mean of sample data
α Scale parameter

τ Translation parameter

m Scaling constant or decomposition level

n_w Translating constant

$\psi*(x)$ complex conjugate function

$\psi(x)$ Mother wavelet

σ Sigmoid function

Variables

bi_c The bias of cell block

bi_f Bias of forget gate

$b_h^{(b)}$ Bias vector for the backward hidden layer

$b_h^{(f)}$ Bias vector for the forward hidden layer

bi_i The bias of input gate

bi_o The bias of output gate

b_o Bias vector for output gate

$\bar{c}_t$ New candidate state

c_t The current cell block in LSTM

f_t The output of the output gate in LSTM

h_t the output of the current state in LSTM

H_t Total hidden vector layer

$\overleftarrow{H}_t$ Hidden vector for backward layer at time t

$\vec{H}_t$ Hidden vector for forward layer at time t

i_t The output of the input gate in LSTM

mm The model counter in the prediction and optimization module

o_t The output of forget gate

O_{fn} The last layer output vector

tt Time interval of signal

$W_{hh}^{(b)}$ Weight vector for output data of backward layer

$W_{hh}^{(f)}$ Weight vector of forward layer output data

Wt_c Weight matrix of the cell block

Wt_f Weight matrix of forget gate

Wt_i Weight matrix of the input gate

W_o Weight vector of the output layer

Wt_o Weight matrix of the output gate

$W_{xhh}^{(b)}$ Weight vector of backward layer input data

$W_{xhh}^{(f)}$ Weight vector for input data of forward layer

x Input variable of dataset

xx The sample data

x_t The current input of LSTM

x_s The solution in optimization

x_s' The solution in optimization

X_t Input data vector at time t

y Target variable of the dataset

Y_k^{ACT} Actual vector for sample k

Y_k^{FOR} Forecasted vector for sample k

Z_s The normalized data with the z-score method

Indices

i Index of variable

i_f Index of function

j Index of variable

k Output layer sample index

t Index of the time sample

1. INTRODUCTION

Increasing housing in urban areas and rapid population growth require cities to adapt to the minimum quality of the people's life. In this regard, smart cities are a viable solution for integrating human capital, public resources, social capital and information, and communication technologies to promote sustainable development [1]. The most important pillars of a smart city include smart energy system, smart community system, smart transportation, and smart health care [2]. Smart energy system means a sustainable energy and energy market management. Energy market participants can adopt their most effective and profitable trading strategies by accurately predicting prices [3]. In a sustainable smart city, the consumers can manage their energy consumption bills using a smart energy management system based on electricity price forecasting (EPF). In addition, various important decisions in competitive electricity markets, such as fuel purchasing, power generation planning, and maintenance planning, can be made based electricity load forecasting (ELF) [4]. In addition, ELF plays an important role in making energy production and consumption decisions in power grids [5]. Therefore, accurate ELF and EPF is an essential requirement for a sustainable energy system and energy market management in a smart city.

Considering the published papers in the literature in recent years, different methods of EPF and ELF can be roughly divided into three categories based on the technique used for forecasting, including (1) statistical methods, (2) machine learning methods [6] (3) hybrid prediction methods [7].

Statistical methods are defined based on the assumptions of each algorithm and add deficiencies to the model [16]. The weakness of statistical models is the inability to accurately predict spike points, especially in EPF models [8]. Linear forecasting models and correlation analysis methods are examples of statistical methods [7]. Linear forecasting methods use a linear function to predict future electricity price and load demand series values. Autoregressive moving average (ARMA), autoregressive (AR), moving average (MA), and autoregressive integrated moving average (ARIMA) models are common instances of linear prediction techniques [9]. [10] have used the ARIMA models to predict locational marginal price (LMP) and a generalized autoregressive conditional heteroskedasticity (GARCH) to improve further the accuracy of EPFs. Correlation analysis methods compute the correlations between variables to match approximate variables in the system and in this way predict future values. Pattern sequence forecasting (PSF) and gray model (GM) are examples of correlation analysis method [7]. [11] have presented an approach to predict the behavior of electricity price and load demand time series using the clustering technique. In this approach, by using clustering, 24-dimensional time series are labeled and these labels are used instead of the actual labels of the samples to predict the future behavior of the time series.

Machine learning models are widely used for EPF and ELF [6]. Also, machine learning methods are used to establish security in smart power grids [12] and measure the costs spent by the dynamic system [13]. Support vector machine (SVM),

extreme learning machine (ELM), and artificial neural networks (ANNs) are the most widely used machine learning methods [14]. Among different ANN models, recurrent neural networks (RNNs) have been widely used in EPF and ELF. RNN models have memory units that can not only learn the features of the current input, but also the past information. This feature is suitable for forecasting tasks [15]. But RNNs cannot capture the long-time dependence of the input during training due to the vanishing gradient problem. Long short-term memory (LSTM) is an exact version of RNN that can analyze long time series by solving the vanishing gradient problem [14]. BiLSTM is a type of LSTM that has a higher accuracy in prediction than LSTM due to having a strong memory to store all past and future features. An accurate forecasting method based on BiLSTM and micro-clustering (MC) for forecasting power system parameters has presented in [16]. To evaluate the performance of the proposed method, in this study, three predictions, including wind speed, load demand, and electricity price, have investigated in different periods using the data set of Ontario province, Canada. The proposed method significantly has improved the prediction accuracy compared to SVM, LSTM, Random Forest (RF), and convolutional neural network (CNN) models, especially at peak points.

Due to the complexity of EPF and ELF environment, researchers use the benefits of hybrid techniques to predict future electricity price and load demand values accurately [17]. Hybrid techniques are a combination of prediction models with methods such as decomposition technique, feature selection method, error correction technique, and optimization method. The decomposition technique is one of the effective methods to increase the model's accuracy in forecasting signals with high fluctuations. Although RNN can map nonlinear features, it cannot learn frequency domain information. To solve this problem, decomposition techniques are used to decompose the original signal into several sub-signals and use a specific learning model to predict each signal [15]. Therefore, the prediction accuracy increases. The wavelet packet transform (WPT) [18], wavelet transform (WT) [19], and empirical mode decomposition (EMD) [20] are examples of decomposition techniques. [21] have presented a forecasting model with LSTM and wavelet technique to predict failure in power system. [22] have used a combination of BiLSTM and WT decomposition technique to predict wind speed. Also, [23] have presented a robust hybrid model that is a combination of a multi-layer BiLSTM network, an LSTM model, a genetic algorithm, and a WT technique to predict future gas consumption. Due to the volatile nature of the electricity market and high fluctuation signals of electricity price and load demand, the combination of WT and BiLSTM significantly improves the accuracy, especially at peak points.

The main requirement in EPF and ELF in the smart city electricity market is to maintain the accuracy of prediction models by reducing the model execution time, but in none of the aforementioned papers the execution time of the proposed model for ELF and EPF is not considered. [24] have used parallel computations to minimize runtime and maintain the accuracy of forecasting models for predicting power grid loads. In this paper, among the SVM, Linear Regression, RF, MLP Regressor, Decision Tree, and Gradient Boosting Regressor models, the decision tree models performed better than the other models, creating a trade-off between model accuracy and execution time. In another study, the forecasting model, which includes preprocessing, forecasting, and optimization modules, is proposed to predict the day-ahead electricity load demand. To predict the load demand of smart grids, a combined ANN has used, which is a multi-model ANN predictor with a supervised architecture and multivariate autoregressive algorithm (MARA) for training. The results of the execution on US DAYTOWN, EKPC, and FE network data show that the proposed model increases the accuracy and reduces the execution time [25]. The common problem with these papers is that different execution times are not considered for a particular model, and only the execution times of different models are compared. [26] have compared different time series techniques for short-term price forecasting (STPF) to optimize battery consumption. ARIMA, Prophet, and XGBoost algorithms have been implemented and compared for STPF. Comparison results represent that the XGBoost model performs better than the ARIMA and Prophet models and has a faster execution time. The execution time of the proposed model in this paper has been determined by automatically adjusting the amount of data for network training, and the effect of other hyperparameters on increasing or decreasing the execution time and accuracy is ignored. Therefore, the important weakness in the papers presented in this paragraph is not investigating the effect of different hyperparameter settings in increasing the accuracy and reducing the execution time for each prediction model and not determining the optimal prediction models based on the accuracy and execution time.

Accurate EPF and ELF is very important in smart city energy management. Consumers can manage their energy consumption by EPF accurately. Also, important decisions in the competitive electricity market are made based on the ELF. As a result, the existence of an accurate model for EPF and ELF is an essential requirement in a smart city. On the other hand, an optimal model for EPF and ELF should have a suitable execution time in addition to high accuracy. The values of hyperparameters of a model play an important role in improving the accuracy and execution time of the model. Therefore, by accurately adjusting these values, optimal models can be obtained for EPF and ELF. On the other hand, all the models obtained by changing the value of hyperparameters of the prediction model are not optimal in terms of accuracy and execution time. As a result, the existence of a suitable optimization algorithm is necessary to determine the optimal models. Pareto optimization is an algorithm that balances different objective functions. Pareto optimization is a suitable algorithm for balancing two objective functions in the electricity market. In this paper, the aim is to present the most optimal model based on the two objective functions of accuracy and execution time, which has the highest prediction accuracy in the shortest time. By implementing the Pareto algorithm on different prediction models obtained from hyperparameter settings, optimal models can be determined based on the two objective functions of accuracy and execution time.

Based on the explanations given and according to the gap in EPF and ELF, the contributions of this paper are as follows:

a) The values of different hyperparameters in the hybrid BiLSTM and LSTM models are modified for each season.

b) The models with different accuracy and execution time are provided in order to use in different situations with varying facilities of hardware and available time in the smart cities.

c) Pareto optimization is utilized for BiLSTM and LSTM models of each season, and optimal models are determined for forecasting to be used in situations with different facilities in the smart cities.

d) The most optimal model that has the best trade-off between accuracy and execution time is determined in the Pareto front charts.

The content of the paper is divided as follows. The method used to predict is described in section 2. Section 3 presents the data set used in the implementation, the metrics for calculating the output error, and the results of the model execution. In section 4, the implementation results are discussed, and the paper's conclusion is stated in section 5.

2. PROPOSED METHODOLOGY

The proposed methodology for EPF and ELF includes four modules, which are described as follows. In the preprocessing module, seasonal data normalization is done. In the decomposition step, the electricity price and load demand target signals are decomposed into four signals by the WT decomposition technique. By substituting each decomposed signal as the target variable of the training dataset, four new datasets are created, and all four datasets enter the feature selection module. Then, by the mutual information (MI) feature selection, for each data set, the degree of similarity between input variables and the target variable is determined, and the variables that their degree of similarity is less than a threshold1 value are deleted. In the second stage of feature selection, based on the interaction gain (IG) method, one of the two input variables in the data set that their degree of similarity to each other is greater than a threshold2 value is removed. With obtaining four data sets for prediction, an LSTM network is defined for each data set in the WT-MI-IG-LSTM model. As a result, in the WT-MI-IG-LSTM model, four LSTM networks are trained to predict four different target signals. Also, for the WT-MI-IG-BILSTM model, four BILSTM networks are trained for prediction. After training the models and performing the prediction, the four obtained prediction vectors are combined to get the final prediction output. Then, the process of fine-tuning the hyperparameters of the models is done to obtain models with diverse and appropriate accuracy and execution time. Finally, the Pareto optimization algorithm is applied to the created models to get Pareto optimal solutions in Pareto front chart. The flowchart of the proposed methodology is shown in Figure 1.

A. Preprocessing module

A.1. Normalization

Due to the sensitivity of machine learning models to the scale of the input data, the normalization operation is used to avoid the effect of data scale on the machine learning models [27]. The method used to normalize the data in this paper is the z-score method. The normalized data set with this method has a mean of 0 and a standard deviation of 1. The formula for normalizing data point xx with the z-score method is as follows:

$$Z\_s = \frac{(xx - \bar{X})}{SS} \quad \textbf{(1)}$$

B. Decomposition module

B.1. Wavelet Transform

WT is an important technique that analyzes a non-stationary signal such as electricity price and load demand series and provides the signal frequency and time associated with that frequency. This technique decomposes signals into approximation and detailed series [20]. WT is divided into two categories: continuous wavelet transform (CWT) and discrete wavelet transform (DWT) [28]. CWT for a signal y(tt) is defined as follows [28]:

$$CWT_y(\alpha, \tau) = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{+\infty} y(tt)\psi(\frac{tt-\tau}{\alpha})ddt \quad \textbf{(2)}$$

CWT values are called wavelet coefficients. Additional information is generated due to CWT creation from continuous scaling of the mother wavelet. Therefore, DWT will be more accurate than CWT when the mother wavelet is scaled and translated using specific scales and positions. DWT has been used in signal analysis recently [20]. DWT is currently obtained by the following formula [28]:

$$DWT_y(m, n\_w) = \alpha_0^{-\frac{m}{2}} \int_{-\infty}^{+\infty} y(tt)\psi * (\alpha_0^{-m} tt - n\_w\tau_0)dtt \quad \textbf{(3)}$$

Before decomposition, the mother wavelet and the number of decomposition levels must be selected, and to choose the most suitable mother wavelet, the characteristics of the mother wavelet and the nature of the signal must be considered [29]. Daubechies wavelets of low order (one to four) are the most suitable wavelets for signals with high fluctuations [30]. Therefore, for decomposition with DWT, the fourth-order Daubechies orthogonal wavelet (db4) is used as the mother wavelet, and the number of decomposition levels is three. As a result, the input signal is decomposed into three different levels A1-A3 (in the approximate series) and D1-D3 (in the detailed series). The three-level decomposition is shown in Figure 2.

Based on the nature of the decomposed signals, A3, D1, D2, and D3 signals are the best signals for EPF and ELF [31].The electricity price signal and its decomposed components for the PJM market in 2006 and 2018 are in Figure 3 and Figure 4, and the electricity load demand signal with its decomposed components for the PJM market in 2006 is shown in Figure 5. Due to the high fluctuations of the electricity load demand signal, the components D1-D3 related to this signal present different patterns. The statistical characteristics of the original electricity price signal and its decomposed components for the PJM market in spring 2006 can be seen in Table 1.

C. Two-stage Feature selection module

There are many inputs, including historical datasets for short-term load forecasting (STLF) and STPF. This data may include inefficient inputs that reduce the accuracy of ELF and EPF. Also, as input features increase, the need for more historical data increases, while the available historical data for prediction is limited. Therefore, the appropriate inputs should be selected using the feature selection method to choose a subset of the best features and filter out other trivial features [31]. The two-stage feature selection method used in this paper is based on MI and IG methods.

C.1. Mutual information

One of the most advanced feature selection techniques is MI [31]. MI is defined based on the concept of entropy. Entropy H(x) for discrete variable x with values x1, x2, ..., xn_t and probabilities $P(x1), P(x2), \ldots, P(xn_t)$ indicates the highest possible uncertainty in variable x and is defined as follows [32]:

$$H(x) = -\sum_{i=1}^{n\_t} P(x_i)\log_2(P(x_i)) \quad \textbf{(4)}$$

The joint entropy H(x,y) represents the total entropy of random variables x and y and is defined by [32]:

$$H(x,y) = -\sum_{i=1}^{n\_t}\sum_{j=1}^{m\_t} P(x_i, y_i)\log_2(P(x_i, y_i)) \quad \textbf{(5)}$$

Table 1. Statistical properties of the electricity price signal for the PJM market in the spring of 2006 and decomposed series

Signal	Mean	Median	Maximum	Minimum	Standard deviation
Original signal	-0.01078	-0.2083	3.723	-1.927	0.9807
A3	-0.01253	-0.03404	2.185	-1.338	0.05789
A2	-0.01018	-0.09789	3.157	-2.362	0.9152
A1	-0.01078	-0.1626	3.562	-1.923	0.9605
D3	0.002348	0.02027	1.976	-1.831	0.709
D2	-0.0006011	-0.02016	1.206	-1.299	0.2895
D1	-1.928e-07	-0.003585	0.9635	-0.9698	0.1984

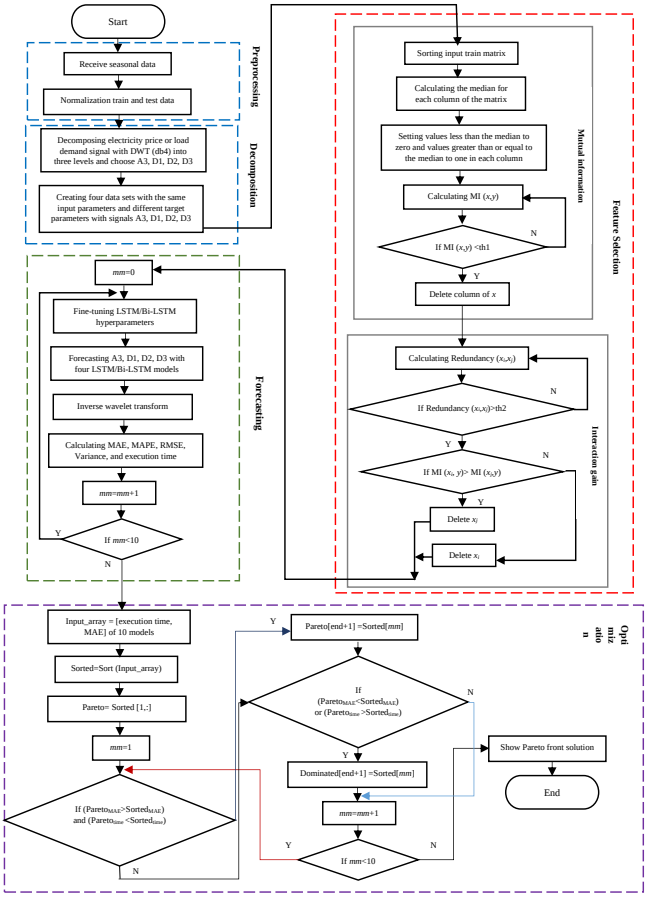


Fig. 1. The flowchart of the proposed method

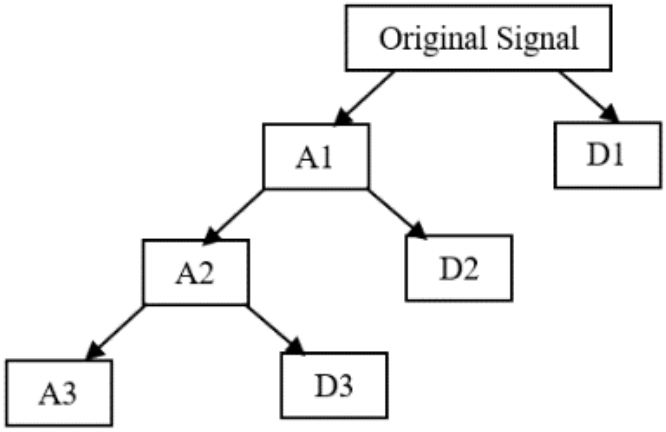


Fig. 2. Three-level decomposition (Original Signal=A3+D1+D2+D3)

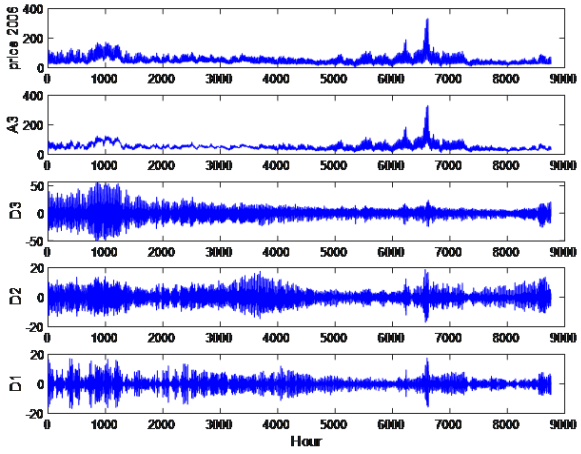


Fig. 3. The electricity price signal and its decomposed components for the PJM market in 2006

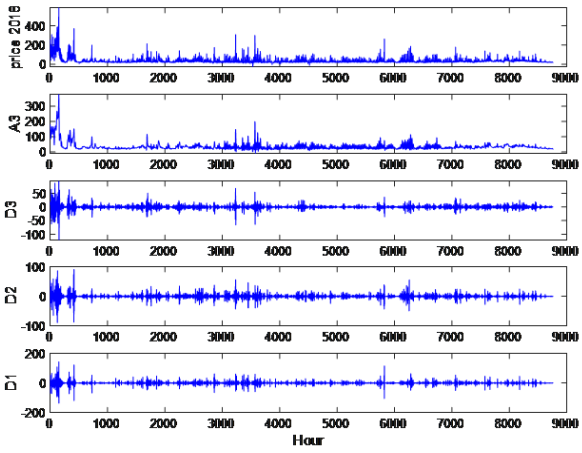


Fig. 4. The electricity price signal and its decomposed components for the PJM market in 2018

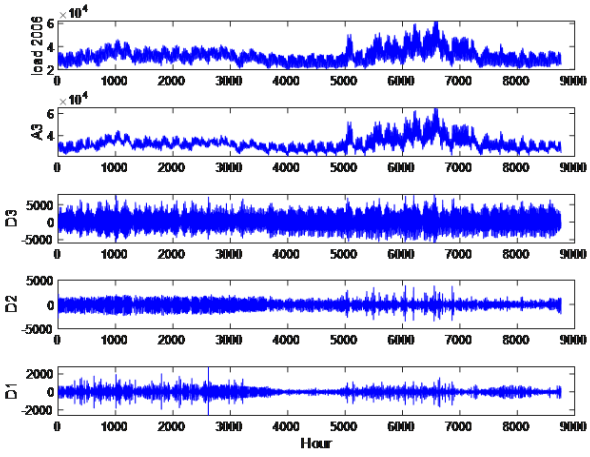


Fig. 5. Electricity load demand signal and its decomposed components for the PJM market in 2006

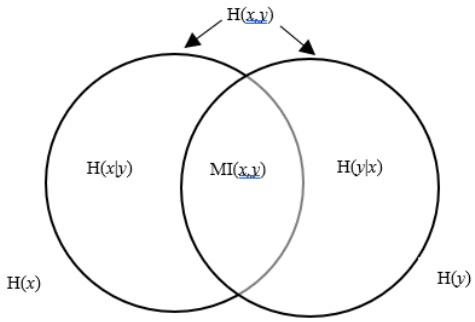


Fig. 6. The relation between mutual information and different entropies

If some certain variables are known and others are unknown, conditional entropy measures the remaining uncertainty [33]:

$$H(\frac{y}{x}) = -\sum_{i=1}^{n\_t}\sum_{j=1}^{m\_t} P(x_i, y_i)\log_2(P(\frac{y_i}{x_i})) \quad \textbf{(6)}$$

The following relation exists between joint entropy and conditional entropy [33]:

$$H(x,y) = H(x) + H(\frac{y}{x}) = H(y) + H(\frac{x}{y}) \quad \textbf{(7)}$$

MI calculates the similarity between output and input. MI(x,y) between two variables is defined as follows [32]:

$$MI(x,y) = \sum_{i=1}^{n\_t}\sum_{j=1}^{m\_t} P(x_i, y_i)\log_2(\frac{P(x_i, y_i)}{P(x_i)P(y_i)}) \quad \textbf{(8)}$$

If two variables are closely related, their MI is high, and vice versa. For two independent variables, the value of MI is zero [32]. The relationship between MI and different entropies is shown in Figure 6.

In the case of STLF and STPF, if the MI value between the input variable (x_i) and the target variable (y) is higher, it means that x_i is a more relevant feature to predict y. To choose a candidate input, the MI between the input and the target is calculated. If the MI value between the input and the target is greater than a threshold1 value, the input feature is selected; otherwise, it is filtered as an irrelevant feature [31].

C.2. Interaction gain

After calculating the value of MI based on equation (7), input attributes are ranked. When two inputs are very alike, one of the input attributes can be removed. In other words, the similarity between the two data inputs with the concept of redundancy is examined. If the similarity of the two specified inputs is greater than a threshold2 value, the input that has a lower correlation with the target variable is removed. As a result, the convergence speed of the learning algorithm increases. For this purpose, the redundancy criterion used for the two candidate inputs, x_i, and x_j, is defined as follows [31]:

$$\text{Re}dundancy(x_i, y_i) = \left|IG(x_i; x_j); y\right| \quad \textbf{(9)}$$

$$IG(x_i; x_j; y) = MI[(x_i, x_j); y] - MI(x_i; y) - MI(x_j; y) \quad \textbf{(10)}$$

After applying the MI and IG methods, a set of hourly electricity price or load demand most related to the target hourly

signal (401st hour) is selected as effective parameters in prediction. Table 2 shows the selected hourly electricity price and load demand for the EPF and ELF of the PJM market in 2006 and 2018.

D. Forecasting module

In the forecasting module, two different models are created with LSTM and BiLSTM networks.

D.1. LSTM network

RNN is a type of deep neural network used in machine translation, speech recognition, and image processing. The main problem of RNN is gradient loss. To resolve this problem, the LSTM network, which is a particular type of RNN and is a combination of short-term and long-term memory, has been presented [34]. The memory cell is the critical component of the LSTM network. LSTM cell structure includes input, output, and forget gate units in the memory block. The input gate memorizes the information of new and previous steps, the forgetting gate removes unimportant information from the memory cell, and the output gate extracts helpful information from the memory cell [6]. The internal calculation flowchart and structure diagram of LSTM at time t are shown in Figure 7 [35].

The basic equations of the LSTM network are as follows [36]:

The basic equations of the LSTM network are as follows (Zhou et al., 2019):

$$i_t = \sigma(Wt_i.[h_{t-1}, x_t] + bi_i) \quad (11)$$

$$f_t = \sigma(Wt_f.[h_{t-1}, x_t] + bi_f) \quad (12)$$

$$c_t = f_t.c_{t-1} + i_t.\bar{c}_t \quad (13)$$

$$\bar{c}_t = \tanh(Wt_c.[h_{t-1}, x_t] + bi_c) \quad (14)$$

$$o_t = \sigma(Wt_o.[h_{t-1}, x_t] + bi_c) \quad (15)$$

$$h_t = o_t.\tanh(c_t) \quad (16)$$

D.2. BiLSTM network

The problem with LSTM is that information that includes future features is ignored [37]. To overcome this shortcoming of LSTM, BiLSTM is proposed. As shown in Figure 8 [38], the general structure of the BiLSTM network has two-way memory. In other words, the BiLSTM network transmits information in two ways: one from the past to the future and the other from the future to the past. In fact, BiLSTM has a powerful memory to store all past and future helpful features with high accuracy [38]. Therefore, the BiLSTM network is more accurate in predicting signals with high random and intermittent behavior because it does not follow the recursive procedure of feedback of past information in an iterative manner that causes error accumulation [39].

In time step t, $X_{(t)}R^{n*r}$ is considered as the mini-batch input data, $H_tRn * n_h$ is the forward hidden state, and the backward hidden state is assumed as $H_tR^{n*n_h}$ for BiLSTM. By integrating the H_t and H_t, the hidden state $H_tR^{n*2n_h}$ is created. Finally, the output data $O_fnR^{n*n_o}$ is computed with the following formulas:

$$\vec{H}_t = \tanh(X_t{W_{xh}}^{(f)} + \vec{H}_{(t-1)})W_{hh}^{(f)} + b_n^{(f)} \quad (17)$$

$$\overleftarrow{H}_t = \tanh(X_t{W_{xh}}^{(b)} + \vec{H}_{(t-1)})W_{hh}^{(b)} + b_n^{(b)} \quad (18)$$

$$O_{fn} = H_tW_o + b_o \quad (19)$$

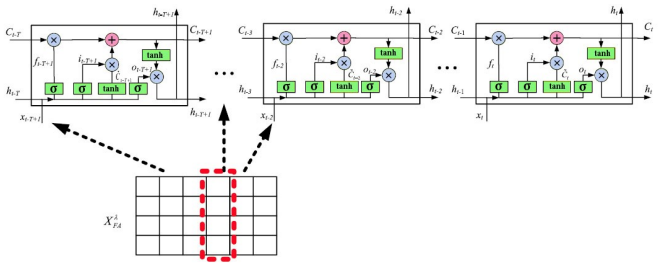


Fig. 7. LSTM internal calculation flowchart and structure diagram at time t

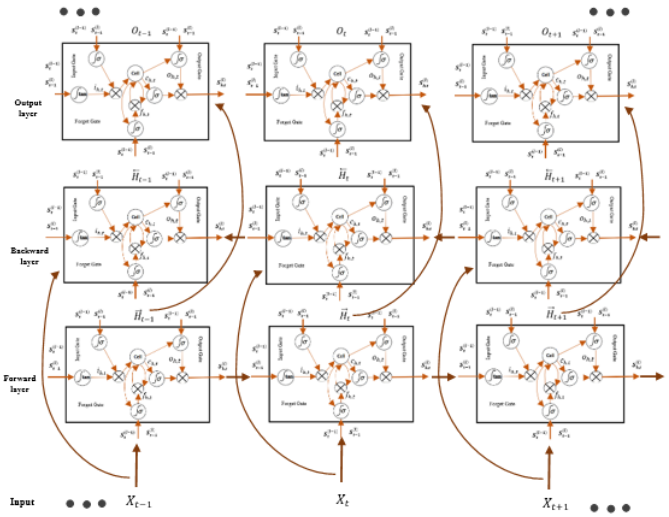


Fig. 8. Structure of BiLSTM network

E. Pareto optimization

Multi-objective optimization (MOO) problems are common in machine learning. Optimizing multiple possibly conflicting objectives is the goal of MOO. MOOs have a set of optimal solutions called a Pareto front, where each point on the Pareto front represents a different trade-off between possibly conflicting objectives [40].

Multi-objective optimization is a setting of optimization with a vector-valued objective function f = (f1, f2, . . ., fm_f): X = X1 × ... × $Xd\ Rm_f$, that each fi_f should be minimized [41].

Generally, the Pareto front is calculated for all conflicting multiple objective trade-offs. A solution $x_s\ \epsilon\ X$ is said to dominate solution $x'_s\ \epsilon\ X$, if x_s is not worse than x'_s in any objective and is absolutely better than x'_s in at least one objective. A solution x_s is called Pareto optimal if and only if it is not dominated by any other solution x'_s. The Pareto front is the collection of all non-dominated solutions [41]. Dominated, non-dominated, and Pareto-front solution sets are shown in Figure 9.

In the framework of hyperparameter tuning defined in this paper, the target model is run ten times with different values of hyperparameters and different structures of LSTM and BiLSTM networks with different numbers of layers and hidden units. The two objective functions of execution time and MAE error are the objective functions of the multi-objective optimization problem. Finally, the Pareto front chart contains the optimal solutions for each model is defined in the specified time interval.

Table 2. The effective parameters for models with the wavelet transform technique

Market		Selected effective features (hours)	
PJM2006		WT-MI-IG-LSTM	WT-MI-IG-BiLSTM
Electricity load	Spring	9,184,189,207,322,358	4,123,136,192,200,202,296,308,315,322
	Summer	2,176,202,236,305,317,382,387,400	4,202,259,305,324
	Fall	145,214,216,240,293,356	67,209,229,292
	Winter	48,211,216,313,322,335,400	46,100,144,173,217,15,320
PJM2006		WT-MI-IG-LSTM	WT-MI-IG-BiLSTM
Electricity load	Spring	243,246,372,388	255,289,292,400
	Summer	118,205,224,262,310,327,392,398	211,222,334,372,377
	Fall	89,343,389,398,400	89,290,318,375,389,390,394,398
	Winter	23,137,250,258,383,396	27,105,238,243,367,299,327,389
PJM2018		WT-MI-IG-LSTM	WT-MI-IG-BiLSTM
Electricity load	Spring	167,247,,382	4,123,136,192,200,202,296,308,315,322
	Summer	158,171,271,277	4,202,259,305,324
	Fall	305,329,383,396	67,209,229,292
	Winter	159,296,308,311,342,355,379,380,383,390,391,392,398,400	46,100,144,173,217,315,320

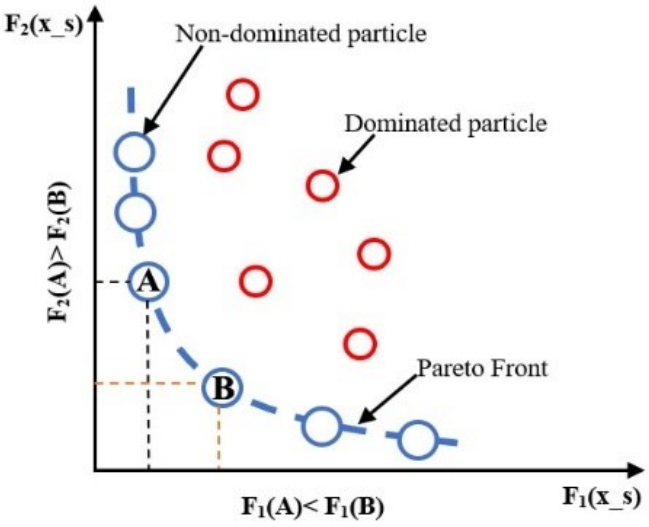


Fig. 9. Dominated, non-dominated and Pareto-front solution set

3. SIMULATION RESULTS

In this paper, an adaptive hybrid model based on WT, feature selection method based on MI and IG, and LSTM or BiLSTM network is used to improve the accuracy of STPF and STLF. To increase the variety of prediction models based on prediction accuracy and model execution time, two different prediction LSTM and BiLSTM models have been created. GradientThreshold, InitialLearnRate and Shuffle hyperparameters are set to appropriate values to create different models with different accuracy and execution time. The optimizer used for the proposed model is the Adam optimizer, the maximum number of training epochs is 500, and the values of Validation frequency and Validation patience are 10 and 5, respectively.

A. Data set

The data used in this paper are related to PJM market electricity prices in 2006 and 2018 and PJM market electricity load demand in 2006. To train the proposed model, the 50-day hourly data of electricity price or load demand in each season have been used as training data. The size of the training matrix is 1200x401, and the last column corresponds to the target parameter. 20% of training data has been used as validation data. Also, the data from one week of each season have been considered as test data. The time interval of the test data for the spring season is from 25 to 31 May, for the summer season from 25 to 31 August, for the autumn season from 24 to 30 November, and for the winter season from 25 to 31 December. The hourly electricity price and load demand data signals of the PJM market in 2006 and 2018 are shown in Figure 10. Simulation results were obtained by running on a PC with an Intel Core i5, 2.71GHz CPU, and 8 GB of RAM. The proposed model is implemented in MATLAB 2021b software.

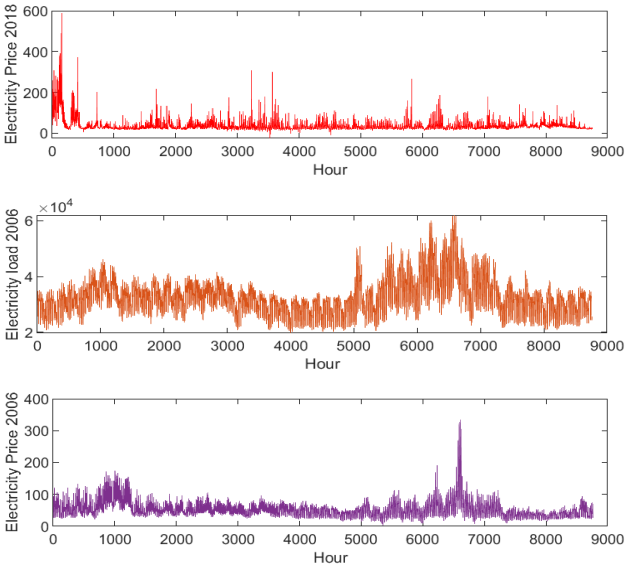


Fig. 10. PJM market hourly electricity price and load demand data in 2006 and 2018

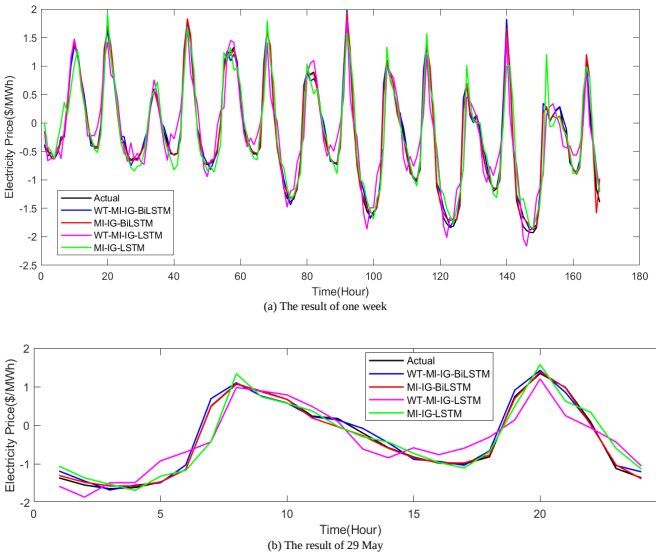


Fig. 11. Forecasted and real electricity price of the spring season in the PJM market in 2006

B. Evaluation metrics

To evaluate the quality of STPF and STLF, five metrics, MAPE, RMSE, MAE, Variance, and R-squared, have been used according to the following formulas:

$$MAE = \frac{1}{nt}\sum_{k=1}^{nt}\left(\left|Y_k^{ACT} - Y_k^{FOR}\right|\right) \tag{20}$$

$$MAPE = \frac{1}{nt}\sum_{k=1}^{nt}\left(\left|\frac{Y_k^{ACT} - Y_k^{FOR}}{Y_k^{ACT}}\right|\right) \tag{21}$$

$$RMSE = \sqrt{\frac{1}{nt}\sum_{k=1}^{nt}\left(\left|Y_k^{ACT} - Y_k^{FOR}\right|^2\right)} \tag{22}$$

$$Variance = \frac{1}{nt}\sum_{k=1}^{nt}\left(\left|\frac{Y_k^{ACT} - Y_k^{FOR}}{Y_k^{ACT}}\right| - MAPE\right)^2 \tag{23}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{nt}\left(Y_k^{FOR} - Y_k^{ACT}\right)^2}{\sum_{k=1}^{nt}\left(Y_k^{FOR} - \bar{Y}_k^{ACT}\right)^2} \tag{24}$$

The MAPE error measure for EPF can be misleading. Because when the electricity price value is very low and close to zero, the MAPE values are very large irrespective of the actual absolute errors, and when electricity price increases, the MAPE values are low [8]. Therefore, although in the comparison of two different models, all the error metrics of one model are lower than the other, it shows the high accuracy of the model, but the comparison of different models with each other in this paper is based on the MAE and RMSE error metrics. The results presented in the tables of this paper are related to the most accurate models obtained.

C. PJM market electricity price forecasting in 2006

One of the largest electricity markets in the world is the PJM electricity market in the United States [31]. The proposed method predicts the electricity price for a week and the next 24 hours in the PJM market in 2006. The results of STPF for the LSTM and BiLSTM networks are presented in Table 4 and Table 5.

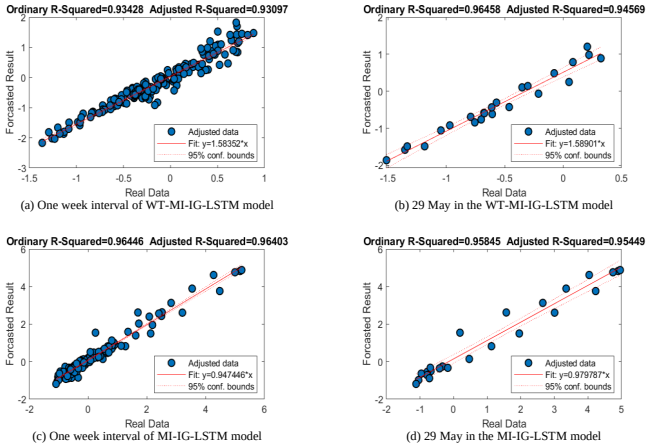
The results of Table 3 show that the LSTM model with the WT technique is better than the LSTM model without the WT technique, except for spring and 29 May. Although the WT-MI-IG-LSTM model is much better than the MI-IG-LSTM model in predicting peak points, but does not predict the valley points well in the period of one week of spring and 29 May. As a result, the WT-MI-IG-LSTM model has a lower prediction accuracy than the MI-IG-LSTM model in these two time periods. In general, the WT-MI-IG-LSTM model shows improvement in prediction accuracy compared to the MI-IG-LSTM model. For example, the values of MAE, MAPE, RMSE, and Variance errors in next week EPF in the summer season for the WT-MI-IG-LSTM model are equal to 0.1118, 0.9962, 0.1474, and 67.1680 with execution time of 123.5469 seconds and for the MI-IG-LSTM model are equal to 0.1341, 1.0591, 0.1679, and 62.6896 with execution time of 51.9306 seconds. The forecasting chart for the spring and 29 May is shown in Figure 11. According to Figure 11, although the MI-IG-LSTM model has less error than the WT-MI-IG-LSTM model, it performs weaker than the WT-MI-IG-LSTM model in predicting peak points, which is due to not using the WT technique. However, according to Figure 12, there is a high correlation between the predicted and actual values in both models in spring, which is due to the excellent performance of the feature selection techniques. The EPF results of BiLSTM models are shown in Table 4. Based on the results of this table, the WT-MI-IG-BiLSTM model has improved accuracy by 17% compared to the best benchmark model, the MI-IG-BiLSTM model in predicting electricity price of the next week of summer, autumn and winter seasons and August 31. For the next week of autumn, due to the MAE value of the WT-MI-IG-BiLSTM model being close to the MI-IG-BiLSTM model and the RMSE criterion of the model with the WT technique being low compared to the model without the WT technique, the WT-MI-IG-BiLSTM model has a more accurate prediction. MI-IG-BiLSTM model is more accurate than WT-MI-IG-BiLSTM for EPF in one week of the spring season and for 29 May, 28 November, and 29 December. The low accuracy of the WT-MI-IG-BiLSTM model in these time

Table 3. EPF results of LSTM-based models for the PJM market in 2006

	WT-MI-IG-LSTM					MI-IG-BIL-LSTM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.2566	0.8881	0.3318	9.6833	250.9748	0.2014	0.9711	0.2729	30.1224	48.9666
29 May	0.3306	0.5529	0.4037	0.3310	-	0.2103	0.4411	0.2786	0.5013	-
The last week of summer	0.1118	0.9962	0.1474	67.1680	123.5469	0.1341	1.0591	0.1679	62.6896	51.9306
31 August	0.1357	0.8860	0.1588	5.4416	-	0.1632	1.1420	0.1854	13.4299	-
The last week of fall	0.1999	0.8390	0.2778	3.5936	184.338	0.2461	0.8066	0.3251	1.9199	18.3058
28 November	0.2232	0.6787	0.3231	1.8049	-	0.2580	0.6699	0.3718	0.8580	-
The last week of Winter	0.2273	0.6431	0.3101	2.4772	152.0808	0.3112	1.1207	0.4566	22.9422	18.5604
29 December	0.1616	0.2308	0.2278	0.1000	-	0.2743	2.7476	0.4020	135.3415	-

Table 4. EPF results of BiLSTM-based models for the PJM market in 2006

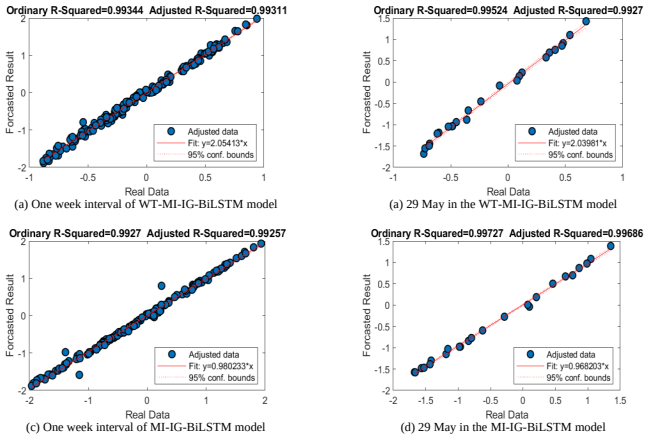
	WT-MI-IG-BIL-STM					MI-IG-BIL-STM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.0791	0.1976	0.0986	0.0898	534.3096	0.0452	0.1392	0.0827	0.1390	141.6562
29 May	0.0919	0.1523	0.1067	0.0236	-	0.0428	0.1351	0.0615	0.0861	-
The last week of summer	0.0447	0.6339	0.0558	42.5349	293.3145	0.0525	1.1165	0.0658	147.0743	47.83
31 August	0.0526	0.1829	0.0701	0.0786	-	0.0719	0.3122	0.0931	0.4149	-
The last week of fall	0.0869	0.4755	0.1167	1.6408	417.2978	0.0859	0.3245	0.1352	0.4660	107.626
28 November	0.1085	0.4740	0.1388	1.0193	-	0.0898	0.2686	0.1139	0.1447	-
The last week of Winter	0.0836	0.2673	0.1109	0.4882	199.9804	0.1127	0.2609	0.1553	0.2596	33.2744
29 December	0.0665	0.2648	0.0869	0.8265	-	0.0603	0.2064	0.0779	0.3806	-



(a) One week interval of WT-MI-IG-LSTM model (b) 29 May in the WT-MI-IG-LSTM model

(c) One week interval of MI-IG-LSTM model (d) 29 May in the MI-IG-LSTM model

Fig. 12. R-Squared values of LSTM models EPF in the spring season of the PJM market in 2006

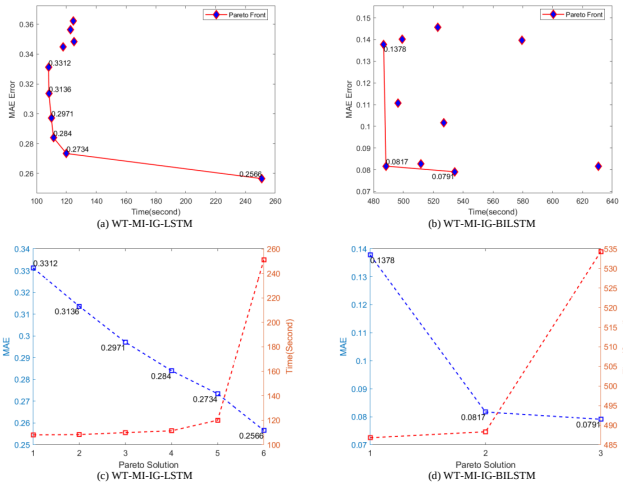


(a) One week interval of WT-MI-IG-BiLSTM model (a) 29 May in the WT-MI-IG-BiLSTM model

(c) One week interval of MI-IG-BiLSTM model (d) 29 May in the MI-IG-BiLSTM model

Fig. 13. R-Squared values of BiLSTM models EPF in the spring season of the PJM market in 2006

intervals is due to the low accuracy of predicting the peak points of the decomposed A3 signal. The forecasting chart for the spring and 29 May in Figure 11 shows the high accuracy in forecasting of MI-IG-BiLSTM model. Also, R-Squared values in Figure 13 prove the high ability of MI and IG techniques. The Pareto chart in Figure 14 shows various examples of the optimal WT-MI-IG-LSTM and WT-MI-IG-BiLSTM models based on different accuracy and execution time, which is the result of tuning the hyperparameters and the structure of the initial models. The optimal models are shown in the Pareto front chart. This chart shows different trade-offs between objective functions in optimal models. According to the chart of the WT-MI-IG-LSTM model in Figure 14, the model with an error of 0.2734 and an execution time of approximately 120 seconds is the closest model to the intersection of the two charts of MAE error and execution time. Therefore, this model has established the best trade-off between accuracy and execution time and among the models presented in the Pareto front chart, this model is the most optimal model. Also, in the chart of the WT-MI-IG-BiLSTM model in Figure 14, the model with an error of 0.0817 and an execution time of approximately 490 seconds is the closest model to the intersection of the two charts and the most optimal model.



(a) WT-MI-IG-LSTM (b) WT-MI-IG-BILSTM

(c) WT-MI-IG-LSTM (d) WT-MI-IG-BILSTM

Fig. 14. EPF Pareto optimization chart for the spring season in the PJM market in 2006

D. PJM market electricity load forecasting in 2006

In this section, the STLF results of the PJM market in 2006 for LSTM and BiLSTM models are given. The results of Table 5 show the higher accuracy of the LSTM model with the WT technique than the LSTM model without the WT technique in all periods except the forecasting of 29 May and 28 November. The sudden jump of the prediction signal of the WT-MI-IG-LSTM model at the peak points of the decomposed signal D3 on 29 May and 28 November has reduced the accuracy of the WT-MI-IG-LSTM model compared to the MI-IG-LSTM. In most cases, LSTM model with WT technique has accuracy improvement compared to LSTM model without WT technique. For example, the values of the MAE, MAPE, RMSE, and Variance error metrics on 29 December for the WT-MI-IG-LSTM technique are 0.1244, 0.3465, 0.1445, and 0.1274, and execution time of 108.3043 seconds, respectively, and for the MI-IG-LSTM technique, they are 0.1280, 0.3107, 0.1924, and 0.2947 and the execution time of 17.5847 seconds. In ELF for one week of winter, due to the MAE value of the WT-MI-IG-LSTM model being close to the MI-IG-LSTM model and the RMSE value of the WT-MI-IG-LSTM model being lower, this model predicts more accurately. The prediction chart for spring and 29 May in Figure 15 and the R-Squared values in Figure 16 have been shown. Table 6 shows the results of BiLSTM models. In this table, the WT-MI-IG-BiLSTM model has 21.8% higher accuracy than the MI-IG-BiLSTM model that is the best benchmark model in all time intervals. For example, the values of MAE, MAPE, RMSE, and Variance errors in the next week ELF in the summer season for the WT-MI-IG-BiLSTM model are equal to 0.0381, 0.1702, 0.0483, 0.9138, and execution time of 321.8649 seconds and for MI-IG-BiLSTM model are equal to 0.0516, 0.3839, 0.0706, 6.0031, and execution time of 103.6589 seconds. In the chart of Figure 15, which shows the ELF of the spring season and 29 May of BiLSTM models, the superiority of the BiLSTM model with the WT technique in forecasting is quite clear. Also, the R-Squared values shown in Figure 17 prove the superiority of the WT-MI-IG-BiLSTM model.

The Pareto chart of the WT-MI-IG-BiLSTM model in Figure 18 shows the optimal models with different accuracy and execution times. The optimal models are the models on the Pareto-front chart. The most optimal model among Pareto front chart models is the model closest to the intersection of two MAE error and ex-

Table 5. ELF results of LSTM-based models for the PJM market in 2006

	WT-MI-IG-STM					MI-IG-STM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.1028	5.4352	0.1352	3286.7566	211.2879	0.1354	5.1887	0.1974	3185.2627	27.5329
29 May	0.1236	314752	0.1427	21992.6719	-	0.0893	1.5076	0.1202	26.0253	-
The last week of summer	0.0877	0.7516	0.1131	33.8942	118.1931	0.1230	1.0195	0.1587	38.2886	34.0933
31 August	0.0770	0.2163	0.0997	0.0971	-	0.1949	0.4212	0.2212	0.1891	-
The last week of fall	0.1243	0.4887	0.1509	5.2385	196.1215	0.1415	0.7422	0.1691	10.9694	18.1792
28 November	0.1404	0.3909	0.1606	0.6637	-	0.1256	1.0449	0.1546	16.1108	-
The last week of Winter	0.1166	0.3083	0.1510	0.9212	108.3043	0.1140	0.2144	0.1698	0.1446	17.5847
29 December	0.1244	0.3465	0.1445	0.1274	-	0.1280	0.3107	0.1924	0.2947	-

Table 6. ELF results of BiLSTM-based models for the PJM market in 2006

	WT-MI-IG-BIL-STM					MI-IG-BIL-STM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.0620	4.4052	0.0779	2203.1555	503.4621	0.0659	3.0856	0.0989	729.6678	184.7813
29 May	0.0578	3.3612	0.0720	210.5173	-	0.0669	7.3528	0.0753	1115.1763	-
The last week of summer	0.0381	0.1702	0.0483	0.9138	321.8649	0.0516	0.3839	0.0706	6.0031	103.6589
31 August	0.0406	0.0849	0.0550	0.0054	-	0.0643	0.1683	0.0932	0.0561	-
The last week of fall	0.0489	0.1331	0.0588	0.1086	389.0196	0.0684	0.3187	0.0781	1.9128	74.0977
28 November	0.0589	0.2365	0.0704	0.5314	-	0.0735	0.4392	0.0786	2.5376	-
The last week of Winter	0.0468	0.0933	0.0639	0.0180	407.8286	0.0701	0.1317	0.1090	0.0498	40.5755
29 December	0.0341	0.0869	0.0424	0.0058	-	0.0345	0.0850	0.0437	0.0098	-

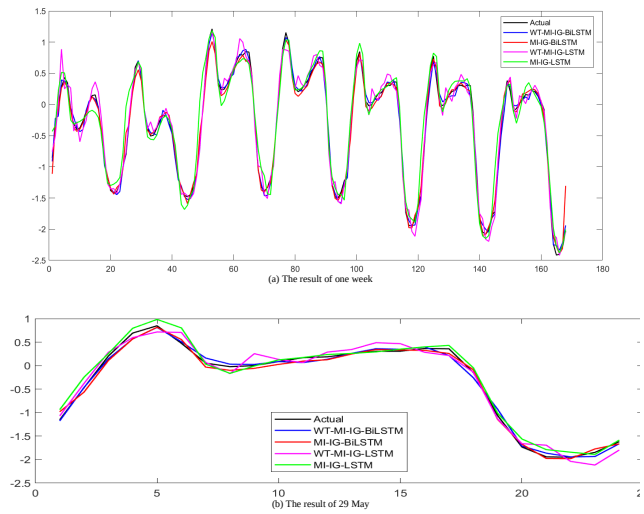


Fig. 15. EPF Pareto optimization chart for the spring season in the PJM market in 2006

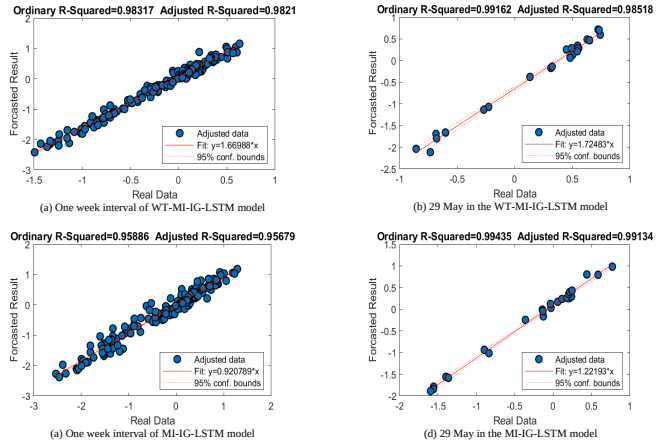


Fig. 16. R-Squared values of LSTM models ELF in the spring season of the PJM market in 2006

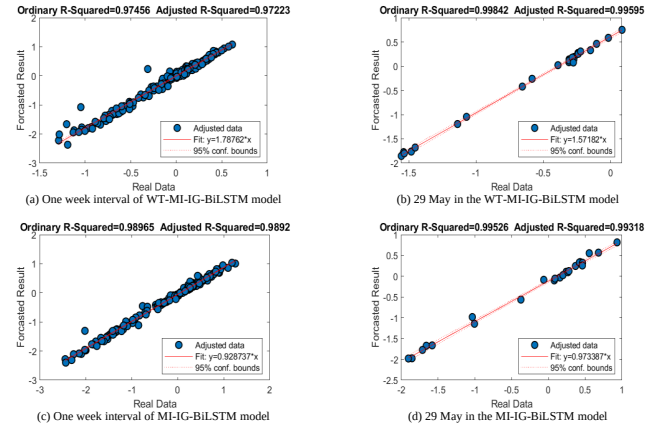


Fig. 17. R-Squared values of BiLSTM models ELF in the spring season of the PJM market in 2006

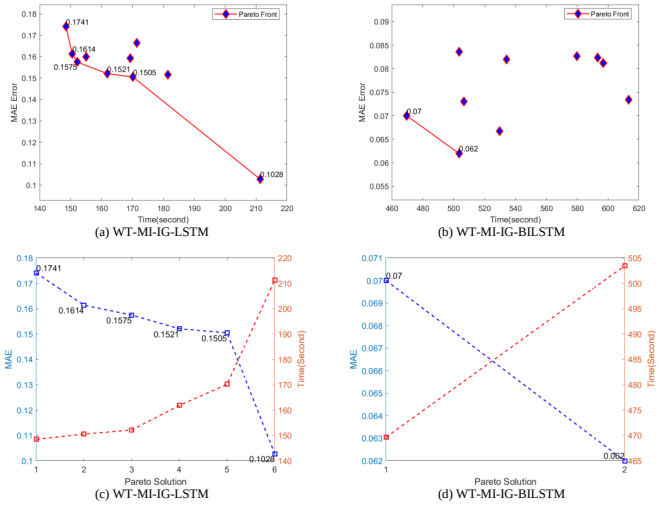


Fig. 18. ELF Pareto optimization chart for the spring season in the PJM market in 2006

ecution time charts. The most optimal model establishes the best trade-off between accuracy and execution time. In the Pareto chart of the WT-MI-IG-LSTM model in Figure 18, the model with an error of 0.1505 and an execution time of approximately 170 seconds is the most optimal model. Also, in the Pareto chart of the WT-MI-IG-BiLSTM model, the model with an error of 0.07 and an execution time of 470 seconds has established a good trade-off between accuracy and execution time and is the most optimal model.

E. PJM market electricity price forecasting in 2018

Unlike in 2006, the PJM market experienced sudden price jumps in 2018. Therefore, STPF in 2018 is very difficult and requires more accurate models. Table 7 shows the STPF results of two different LSTM models. The results of this table show the higher accuracy of the WT-MI-IG-LSTM model than the MI-IG-LSTM model in all time intervals and using the WT technique has improved the accuracy in the LSTM model. The chart of the spring season and 29 May for two models in Figure 19 shows the higher prediction accuracy of WT-MI-IG-LSTM, especially at the peak points. For the EPF of one week in the spring season, to increase the prediction accuracy of LSTM, the A3 decomposed signal has been used instead of the original signal for prediction.

The use of the A3 signal is due to the high similarity of this signal to the original signal. The diagram in Figure 20 also shows the high R-Squared values of the spring season for the WT-MI-IG-LSTM model. The results of BiLSTM models are shown in Table 8. This table show a superiority of 18.16% in the EPF accuracy of the WT-MI-IG-BiLSTM model in all periods except 29 May and 31 August. 29 May and 31 August 2018 have the highest jump in electricity prices compared to all seasons of the data set. On the two dates mentioned, the WT-MI-IG-BiLSTM model has not accurately predicted the peak point of the decomposed signal D1, which has the highest frequency among the decomposed signals. As a result, the accuracy of the WT-MI-IG-BiLSTM model is lower than the MI-IG-BiLSTM model on 29 May and 31 August.

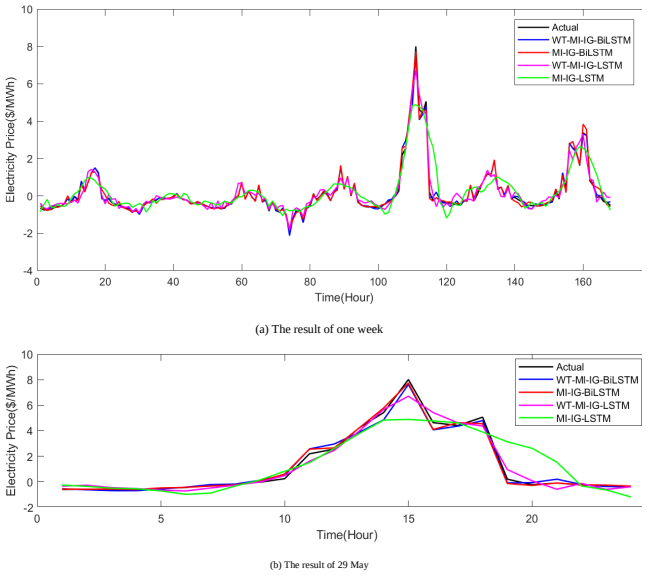
The prediction chart of the spring season and 29 May in Figure 19 shows the superiority of the BiLSTM model without the WT technique compared to the BiLSTM model with the WT technique. Also, the high values of R-Squared in Figure 21 also

Table 7. EPF results of LSTM-based models for the PJM market in 2018

	WT-MI-IG-STM					MI-IG-STM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.2246	1.9826	0.3210	270.7707	164.7424	0.3925	1.6070	0.6055	10.6058	20.9344
29 May	0.3299	0.7308	0.4432	1.3484	-	0.7511	2.6131	1.1833	22.6246	-
The last week of summer	0.4168	3.2491	0.7297	383.968	76.744	0.6380	2.0455	1.4418	49.1786	19.8343
31 August	0.9017	12.8778	1.5456	2451.7900	-	0.9959	4.5509	2.6282	242.1696	-
The last week of fall	0.2447	1.6602	0.3587	24.5946	77.2159	0.4589	1.8406	0.17670.7132	20.5981	29.9289
28 November	0.2511	1.8767	0.3587	41.0075	-	0.6183	2.5679	0.9292	40.7006	-
The last week of Winter	0.1117	0.2138	0.1521	0.1260	71.9771	0.2111	0.3447	0.2751	0.1704	24.7726
29 December	0.0910	0.2712	0.1110	0.1403	-	0.2164	0.4862	0.2765	0.5484	-

Table 8. EPF results of BiLSTM-based models for the PJM market in 2018

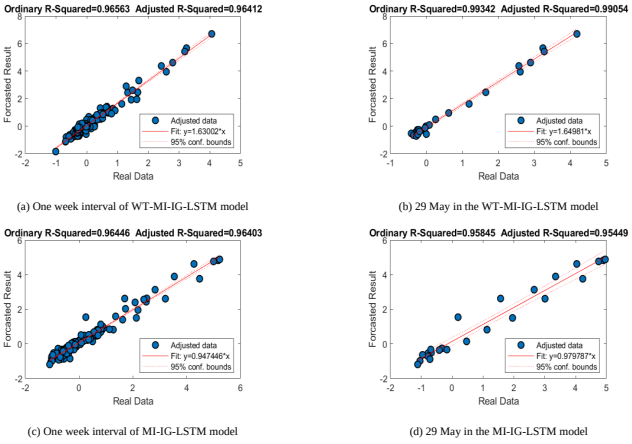
W	WT-MI-IG-BIL-STM					MI-IG-BIL-STM				
	MAE	MAPE	RMSE	VAR	TIME(S)	MAE	MAPE	RMSE	VAR	TIME(S)
The last week of spring	0.0685	0.3364	0.1179	1.0801	576.6634	0.0883	0.4199	0.1437	1.5342	132.7026
29 May	0.1973	0.5822	0.2611	1.4414	-	0.1463	0.3133	0.2245	0.3984	-
The last week of summer	0.1491	1.3176	0.3788	103.3688	383.95	0.1431	2.1976	0.4045	560.7831	177.0194
31 August	0.3796	6.3995	0.7460	685.0602	-	0.3642	13.6588	0.9186	3767.1221	-
The last week of fall	0.0919	0.5030	0.1291	2.2659	354.1735	0.1076	0.3477	0.1767	0.5167	147.8475
28 November	0.1136	0.5631	0.1532	3.6315	-	0.1166	0.1185	0.2200	0.0127	-
The last week of Winter	0.0542	0.1000	0.0720	0.0222	524.3325	0.0975	0.1702	0.1233	0.0493	149.5073
29 December	0.0476	0.1160	0.0657	0.0278	-	0.0879	0.2150	0.0990	0.0547	-



(a) The result of one week

(b) The result of 29 May

Fig. 19. Forecasted and real electricity price of the spring season in the PJM market in 2018



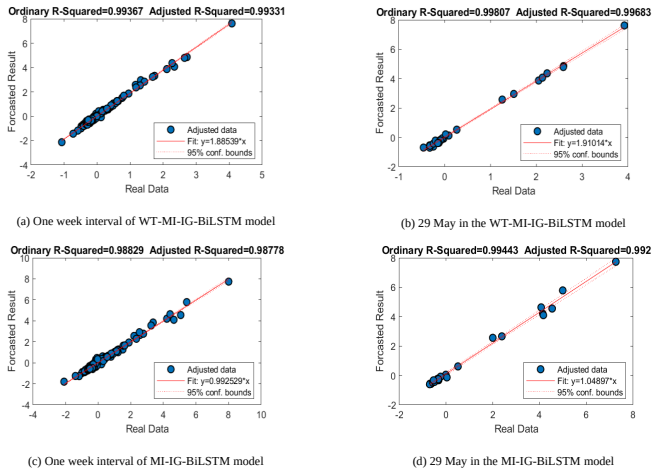
(a) One week interval of WT-MI-IG-LSTM model

(b) 29 May in the WT-MI-IG-LSTM model

(c) One week interval of MI-IG-LSTM model

(d) 29 May in the MI-IG-LSTM model

Fig. 20. R-Squared values of LSTM models EPF in the spring season of the PJM market in 2018



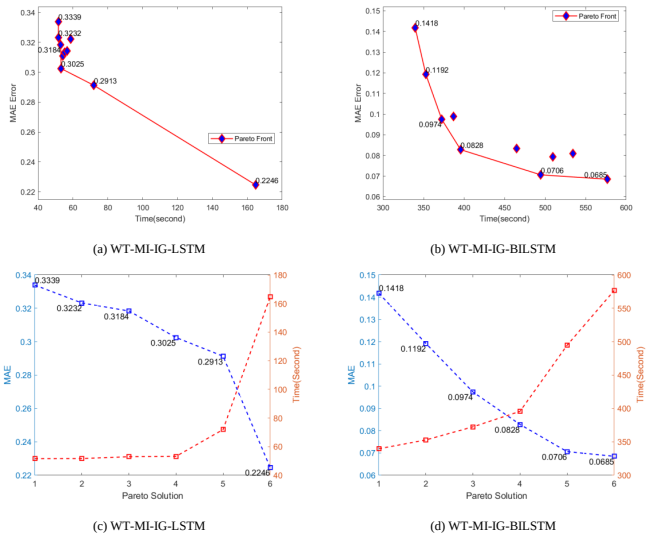
(a) One week interval of WT-MI-IG-BiLSTM model

(b) 29 May in the WT-MI-IG-BiLSTM model

(c) One week interval of MI-IG-BiLSTM model

(d) 29 May in the MI-IG-BiLSTM model

Fig. 21. R-Squared values of BiLSTM models EPF in the spring season of the PJM market in 2018



(a) WT-MI-IG-LSTM

(b) WT-MI-IG-BILSTM

(c) WT-MI-IG-LSTM

(d) WT-MI-IG-BILSTM

Fig. 22. EPF Pareto optimization chart for the spring season in the PJM market in 2018

indicate the high correlation between the predicted and actual values in both BiLSTM models.

In the Pareto chart of Figure 22, different WT-MI-IG-LSTM and WT-MI-IG-BiLSTM models with different accuracies and execution times are presented. Among the optimal models presented in the Pareto front chart, the model that is closer to the intersection of the two MAE and execution time charts is the most optimal model. In the WT-MI-IG-LSTM chart, the most optimal model is the model with an error of 0.2913 and an execution time of 70 seconds. Also, the most optimal model of the WT-MI-IG-BiLSTM Pareto chart is the model with an error of 0.0828 and an execution time of approximately 400 seconds, which has established a good trade-off between accuracy and execution time.

4. DISCUSSION

In this paper, the adaptive hybrid WT-MI-IG-BiLSTM model for STPF and STLF is implemented on PJM market data in 2006 and

2018. Also, WT-MI-IG-LSTM, MI-IG-LSTM and MI-IG-BiLSTM models have been implemented in order to confirm the high accuracy of the proposed model. BiLSTM model is more accurate than LSTM in forecasting due to having a strong memory to store all past and future features. WT technique decomposes the original electricity price and load demand signals into signals with different frequencies. Adding this technique to the original signal prediction models makes it possible to use a suitable prediction model with hyperparameter settings and different number of hidden units based on the fluctuations of decomposed signals. Therefore, more complex models can be considered for signals with high fluctuations and simpler models for signals with low fluctuations, and models with high accuracy and suitable execution time can be obtained for EPF and ELF. While in forecasting models without WT technique, it is not possible to determine different models for components with different frequencies in the signal, and the selection of the forecasting model is based on the original electricity price and load demand signals and not on the decomposed signals. The results of all predictions of the proposed models are presented in the tables and the charts based on the most accurate results. Based on the results, the use of WT technique increases the prediction accuracy of the models. By decomposing the original signal by the WT technique into signals with different frequencies, the most important issue is to choose a suitable prediction model for each signal. If complex models are considered for prediction of all decomposed signals, with the increasing prediction accuracy, the model execution time also increases. Also, if simple models are used to predict signals in order to reduce the execution time, the accuracy of the prediction will be greatly reduced. Therefore, a proper trade-off between accuracy and execution time should be established with the appropriate settings of the hyperparameters of the prediction model and the determination of the appropriate number of hidden units in the model. Considering using WT technique in WT-MI-IG-BiLSTM and WT-MI-IGLSTM models, proper settings of hyperparameters have performed on these models. For example, in the Pareto plot of STLF with BiLSTM for the PJM market in 2018, shown in Figure 22, four BiLSTM models have utilized to achieve MAE error of 0.0828. First and second BiLSTM networks have four layers of input, bilstm, fully connected, and regression and the other two BiLSTM networks have five layers of input, bilstm, dropout, fully connected, and regression. The number of hidden units of the bilstm layer in first and forth models is equal to 90, for the second model, it is equal to 75, and in the third model, it is equal to 82. The value of the gradient threshold for all networks is one. The value of initial learning rate hyperparameters for the first, third, and fourth networks is 0.01, and for the second network is 0.009. Also, the probability value of the dropout layer for the third and fourth BiLSTM network is equal to 0.5. The same process is followed to create other models with different MAE error values and execution times. After performing various settings on the hyperparameters and the prediction model architecture, to determine the optimal prediction models, the Pareto optimization algorithm was implemented and the optimal models were identified in the Pareto front chart. The models that are in the Pareto front chart are optimal in one of the two objective functions of accuracy and execution time. In other words, the first model on the Pareto front, which is the leftmost model, has the highest error rate, but it also has a shorter execution time. Because this model is made of simple prediction models, which causes low accuracy of the model. Therefore, the first model on the Pareto front has optimal execution time but low accuracy. Gradually, as we move from the first model to the last model (the rightmost model) of the Pareto front, the error decreases and the execution time increases. Finally, the last Pareto model has the lowest error and the longest execution time. The most optimal model in terms of accuracy and execution time should have a proper trade-off between accuracy and execution time. The model closest to the intersection of two MAE error charts and execution time in Pareto charts is the most optimal model in terms of accuracy and execution time. The most optimal model has the best trade-off between accuracy and execution time.

5. CONCLUSIONS

Proper management of energy consumption and commerce in the energy market is very important in a smart city. So, accurate STPF and STLF are critical for market participants in deregulated electricity markets. In this paper, a hybrid and adaptive model based on mutual information (MI) and interaction gain (IG) feature selection methods and wavelet transform (WT) with BiLSTM networks called WT-MI-IG-BiLSTM was implemented. By using the WT technique in the proposed method, the weekly data of electricity price and load demand decomposed into signals with different frequencies. Then the appropriate model for electricity price forecasting (EPF) and electricity load forecasting (ELF) was determined based on the fluctuations of the decomposed signal. Therefore, for decomposed signals with high fluctuations, more complex prediction models were determined and for signals with low fluctuations, simpler prediction models were determined. This diversity in the selection of prediction models based on signal fluctuations shows the high power of WT technique to create models with high accuracy and suitable execution time. Based on the results, the WT-MI-IG-BiLSTM model has improved accuracy by 17-18.16% in EPF and 21.8% in ELF compared to the MI-IG-BiLSTM model, which is the best benchmark model. On the other hand, with different settings of hyperparameters of each original signal prediction model and obtaining different models with different execution time and accuracy, it becomes necessary to choose optimal model for EPF and ELF in smart cities. Pareto optimization algorithm plays the main role in choosing the optimal prediction model. By implementing this algorithm, based on two MAE error and model execution time objective functions, non-dominated solutions have provided in the Pareto front as optimal models for both WT-MI-IG-LSTM and WT-MI-IG-BiLSTM models. Finally, the model closest to the intersection of two MAE error and execution time charts in Pareto front charts, which establishes the best trade-off between accuracy and execution time, was chosen as the most optimal model for EPF and ELF. In addition, it is possible to use other models in the Pareto front chart for forecasting based on the required forecasting accuracy, time and available computing resources.

REFERENCES

1. de Guimarães JCF, Severo EA, Felix Júnior LA, da Costa WPLB, Salmoria FT. Governance and quality of life in smart cities: Towards sustainable development goals. J Clean Prod. 2020 Apr 20;253.

2. Silva BN, Khan M, Han K. Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. Vol. 38, Sustainable Cities and Society. Elsevier Ltd; 2018. p. 697–713.

3. Mujeeb S, Javaid N, Ilahi M, Wadud Z, Ishmanov F, Afzal MK. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. Sustainability (Switzerland). 2019 Feb 14;11(4).

4. Elattar EE, Goulermas J, Wu QH. Electric load forecasting based on locally weighted support vector regression. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews. 2010 Jul;40(4):438–47.

5. Kumar S, Hussain L, Banarjee S, Reza M, Tech B, Students Y. Energy Load Forecasting using Deep Learning Approach-LSTM and GRU in Spark Cluster.

6. Tschora L, Pierre E, Plantevit M, Robardet C. Electricity price forecasting on the day-ahead market using machine learning. Appl Energy. 2022 May 1;313.

7. Kong X, Li C, Wang C, Zhang Y, Zhang J. Short-term electrical load forecasting based on error correction using dynamic mode decomposition. Appl Energy. 2020 Mar 1;261.

8. Weron R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. Vol. 30, International Journal of Forecasting. Elsevier B.V.; 2014. p. 1030–81.

9. Soliman A hady S, AI-Kandari AM. Electrical load forecasting: modeling and model construction. ELSEVIER; 2010.

10. Zhao Z, Wang C, Nokleby M, Miller CJ. Improving Short-Term Electricity Price Forecasting Using Day-Ahead LMP with ARIMA Models. 2017.

11. Martínez-Alvarez F, Troncoso A, Riquelme JJ, Jesu´s J, Aguilar-Ruiz S. Energy Time Series Forecasting Based on Pattern Sequence Similarity. 2010.

12. Moradi M, Weng Y, Lai YC. Defending Smart Electrical Power Grids against Cyberattacks with Deep Q -Learning . PRX Energy. 2022 Nov 28;1(3).

13. Razavi SE, Moradi MA, Shamaghdari S, Menhaj MB. Adaptive optimal control of unknown discrete-time linear systems with guaranteed prescribed degree of stability using reinforcement learning. Int J Dyn Control. 2022 Jun 1;10(3):870–8.

14. Peng L, Liu S, Liu R, Wang L. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy. 2018 Nov 1;162:1301–14.

15. XY Zhang, S Kuenzel, N Colombo, C Watkins. Short-Term Load Forecasting Method based on. 2021;

16. Jahangir H, Tayarani H, Gougheri SS, Golkar MA, Ahmadian A, Elkamel A. Deep Learning-Based Forecasting Approach in Smart Grids with Microclustering and Bidirectional LSTM Network. IEEE Transactions on Industrial Electronics. 2021 Sep 1;68(9):8298–309.

17. Tian C, Hao Y, Hu J. A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization. Appl Energy. 2018 Dec 1;231:301–19.

18. Shayeghi H, Ghasemi A, Moradzadeh M, Nooshyar M. Simultaneous day-ahead forecasting of electricity price and load in smart grids. Energy Convers Manag. 2015 May 1;95:371–84.

19. Zhang CY, Chen CLP, Gan M, Chen L. Predictive Deep Boltzmann Machine for Multiperiod Wind Speed Forecasting. IEEE Trans Sustain Energy. 2015 Oct 1;6(4):1416–25.

20. Zhang Y, Li C, Li L. Wavelet transform and Kernel-based extreme learning machine for electricity price forecasting. Energy Systems. 2018 Feb 1;9(1):113–34.

21. Branco NW, Cavalca MSM, Stefenon SF, Leithardt VRQ. Wavelet LSTM for Fault Forecasting in Electrical Power Grids. Sensors. 2022 Nov 1;22(21).

22. Xiang J, Qiu Z, Hao Q, Cao H. Multi-time scale wind speed prediction based on WT-bi-LSTM. 2020; Available from: https://doi.org/10.1051/matecconf/202030

23. Su H, Zio E, Zhang J, Xu M, Li X, Zhang Z. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. 2019;178:585–97. Available from: https://hal.archives-ouvertes.fr/hal-02428545

24. Zainab A, Syed D, Ghrayeb A, Abu-Rub H, Refaat SS, Houchati M, et al. A multiprocessing-based sensitivity analysis of machine learning algorithms for load forecasting of electric power distribution system. IEEE Access. 2021;9:31684–94.

25. Ahmad A, Javaid N, Mateen A, Awais M, Khan ZA. Short-Term load forecasting in smart grids: An intelligent modular approach. Energies (Basel). 2019 Jan 1;12(1).

26. Preniqi V, Mishra BK, Thakker D, Feigl E, Mokryani G, Abdullatif A, et al. Comparative Study of Shortterm Electricity Price Forecasting Models to Optimise Battery Consumption. In: 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics) [Internet]. IEEE; 2020. p. 342–9. Available from: https://ieeexplore.ieee.org/document/9291599/

27. Bouktif S, Fiaz A, Ouni A, Serhani MA. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies (Basel). 2018;11(7).

28. Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. Energy. 2019 Nov 15;187.

29. SK Aggarwal, LM Saini, A Kumar. Electricity Price Forecasting in Ontario Electricity Market Using Wavelet Transform in Artificial Neural Network Based Model. 2008;

30. da Rocha Reis AJ, Alves da Silva AP. Feature extraction via multiresolution analysis for short-term load forecasting. IEEE Transactions on Power Systems. 2005 Feb;20(1):189–98.

31. Memarzadeh G, Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm. Electric Power Systems Research. 2021 Mar 1;192.

32. Amjady N, Keynia F. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. IEEE Transactions on Power Systems. 2009;24(1):306–18.

33. Kwak N, Choi CH. Input Feature Selection by Mutual Information Based on Parzen Window. 2002.

34. Liu H, Mi X wei, Li Y fei. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manag. 2018 Jan 15;156:498–514.

35. Meng A, Wang P, Zhai G, Zeng C, Chen S, Yang X, et al. Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization. Energy. 2022 Sep 1;254.

36. Zhou S, Zhou L, Mao M, Tai HM, Wan Y. An Optimized Heterogeneous Structure LSTM Network for Electricity Price Forecasting. IEEE Access. 2019;7:108161–73.

37. Toubeau JF, Bottieau J, Vallee F, de Greve Z. Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets. IEEE Transactions on Power Systems. 2019 Mar 1;34(2):1203–15.

38. Huang CG, Huang HZ, Li YF. A Bidirectional LSTM Prognostics Method Under Multiple Operational Conditions. IEEE Transactions on Industrial Electronics. 2019 Nov 1;66(11):8792–802.

39. Tong W, Li L, Zhou X, Hamilton A, Zhang K. Deep learning PM2.5 concentrations with bidirectional LSTM RNN. Air Qual Atmos Health. 2019 Apr 2;12(4):411–23.

40. Navon A, Shamsian A, Chechik G, Fetaya E. Learning the Pareto Front with Hypernetworks. 2020 Oct 8; Available from: http://arxiv.org/abs/2010.04104

41. Horn D, Bischl B. Multi-objective parameter configuration of machine learning algorithms using model-based optimization. 2016.