IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 September 2023, accepted 14 September 2023, date of publication 18 September 2023,
date of current version 28 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3316516

RESEARCH ARTICLE

Short-Term Power Load Forecasting in FGSM-Bi-LSTM Networks Based on Empirical Wavelet Transform

QINGCHAN LIU[1], JIANING CAO[2,3], JINGCHENG ZHANG[4], YAO ZHONG[1], TINGJIE BA[1], AND YIMING ZHANG[1]

[1]Measurement Center, Yunnan Power Grid Company Ltd., Kunming 650051, China
[2]Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming, Yunnan 650500, China
[3]Kunming Holy Intelligent Technology Company Ltd., Kunming, Yunnan 650000, China
[4]Faculty of Science, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

Corresponding author: Jingcheng Zhang (2431487652@qq.com)

This work was supported by the Science and Technology Project of China Southern Power Grid Company Ltd. under Grant YNKJXM20210147 and Grant YNKJXM20220010.

ABSTRACT In this paper, a prediction model based on Empirical Wavelet Transform (EWT) for FGSM-Bi-LSTM network is proposed to address the short-term power load forecasting problem. The model performs noise reduction on the data by combining time windows with EWT. Additionally, the model stability is enhanced by introducing the Fast Gradient Sign Method (FGSM) to generate adversarial samples. Finally, case experiments based on real-world power station load data are conducted. The results demonstrate that compared with prediction models such as LSTM, ARIMA, XGBoost, QR-GRU, and Transformer, the proposed FGSM-Bi-LSTM model reduces RMSE by 85.22%, improves MAE by 62.60%, and increases r by 9.83%. This demonstrates the strong generalization ability of the model to be extended to other time series prediction tasks.

INDEX TERMS Bi-LSTM, EWT, deep neural networks, FGSM, short-term load forecast, smart grid.

I. INTRODUCTION

With the continuous development of power systems and the advancement of smartification, transformer load forecasting has become a key topic in the power field. Accurate load prediction is crucial for power system operation and energy management, which helps to optimize the allocation of power resources, improve the quality of power supply, as well as achieve energy saving and emission reduction [1]. However, due to the complex dynamics of power loads, current prediction algorithms generally suffer from low accuracy [2]. Lower prediction accuracy may lead to waste of resources or even safety hazards due to equipment overload on the one hand, while on the other hand, it may lead to problems such as power outages, which may bring serious impacts on the stable operation of the society [3]. Given the highly dynamic and periodic nature of power load data, current research on load forecasting is mainly divided into short-term, medium-term and long-term categories [4]. Among them, short-term load forecasting receives the most attention from decision makers, because it directly supports power supply authorities in the regulation and energy management of smart grids.

In recent years, with more and more methods being applied to short-term load forecasting, the error rate of prediction has been reduced and the accuracy of power load forecasting has been improved. Erdogdu combined autoregressive integrated moving average (ARIMA) with cointegration analysis for power load analysis [5]. However, for short-term time series with complex seasonal patterns, ARIMA does not capture seasonal variation well. Compared to traditional time-series prediction techniques such as ARIMA, various types of machine learning models have better learning effect on the complex dynamic changes of power load. Models such as Artificial Neural Networks (ANN) [6], XGBoost (eXtreme Gradient Boosting) [7], and Support Vector Machine (SVM) [8] have also gained widespread use in

The associate editor coordinating the review of this manuscript and approving it for publication was Vlad Diaconita.

power load forecasting. Further, with the increase in computational power, deep learning was applied to time series prediction. Where, Shi et al. applied RNN (recurrent neural network) to household electricity load prediction, although better results were obtained, but the effect is lacking for sample data with high feature complexity [9]. Niu et al. utilized Gated Recurrent Units (GRU) for wind power generation forecasting [10]. Ageng et al. proposed an hourly load forecasting framework combining data preparation and Long Short-Term Memory (LSTM) [11]. Atef et al. on the other hand introduced a deep Bi-LSTM (bi-directional long-short term memory) approach to enhance the accuracy of residential electricity demand forecasting [12]. It is worth noting that LSTM and GRU are able to handle the temporal characteristics of power load data well due to the RNN-based structure that establishes links between neurons in the hidden layer. However, they are not accurate enough in predicting sequences with the presence of high frequency fluctuations and large noise. Therefore, effective noise removal techniques need to be designed to improve the robustness of the model.

Further, in order to overcome the lack of accuracy and narrow applicability of a single model in power load prediction modeling, combined prediction approaches have emerged. Li et al. used sample entropy (SE) to reconstruct the sequence after complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) decomposition, and proposed a CEEMDAN-SE-LSTM model combined with LSTM. And the accuracy of the model is proved by example validation [13]. Jha et al. combined LSTM with a random forest approach [14]. On the other hand, Zhang et al. designed a multi-objective optimization algorithm to balance stability and accuracy, partially overcoming the shortcomings of individual models [15]. However, the complexity of the above models also leads to high computational costs and is prone to overfitting when training data is limited. This results in unstable predictions in the short term. Therefore, there is a need for a new model that can effectively denoise and enhance robustness, while avoiding reduced computational efficiency due to high complexity.

As an important component in time series forecasting, noise processing has received extensive research attention. In order to reduce noise in time series data, scholars have utilized signal processing techniques for data denoising [16], [17], [18]. For instance, Laouafi et al. developed a wavelet decomposition method that improved the accuracy of power load forecasting [19]. Luo et al. introduced empirical modal decomposition (EMD) to reduce interference signals in their study of the distribution network situational awareness problem [20]. Dong then proposed a short-term load prediction method by decomposing the original load sequence using EEMD (ensemble empirical mode decomposition) [21]. While employing signal processing techniques such as EMD, it's important to be aware of the potential challenge of mode mixing during the decomposition process. Mode mixing occurs when distinct components of the signal become entangled or mixed together, making it difficult to separate them accurately. This phenomenon can result in a reduction in the accuracy of the decomposition results and can pose challenges in interpreting and effectively utilizing the decomposed components [22]. For this reason, this paper combines EWT with sliding window to decompose the load of smart grid, which ensures the accuracy of the data while reducing the noise.

Although an increasing number of machine learning (ML) studies have been applied to power systems, research shows that data-driven event causality analysis can be vulnerable to well-crafted malicious data, which even state-of-the-art smart meters struggle to detect [23]. Chen et al. demonstrated that most ML algorithms proposed in power systems are vulnerable to adversarial examples, resulting in unpredictable harm to the system [24]. Given this potential vulnerability, securing load forecasting models against adversarial attacks is an important issue to address. To guard against such adversarial attacks, researchers have proposed various strategies for generating adversarial samples to enhance the robustness of the model [25], [26], [27], [28]. However, the iterative process of these methods is more cumbersome. Whereas FGSM is used as an adversarial machine learning technique, especially in deep learning models, for generating adversarial samples and is widely used in image recognition tasks to evaluate the robustness of models [29]. But it is rarely applied in time series forecasting [30]. It was not until 2019 that Fawaz et al. first (to be sure) applied adversarial training to the task of classifying time series, demonstrating that generating adversarial samples for training of the FGSM would dramatically improve the robustness of the model [31]. Based on that, this paper introduces FGSM to the task of load time series prediction in power systems for the first time to improve robustness against potential attacks.

In summary to address the shortcomings of current research, this paper proposes a prediction model based on EWT and FGSM-Bi-LSTM network as shown in Fig. 1, and by example verifying the robustness of the model. The main research contributions of this paper are as follows:

1. EWT is combined with time windows to decompose the original data forwardly, reducing noise and avoiding data loss.
2. FGSM is first introduced into power system time series forecasting. The loss function integrating regression and classification errors improves robustness.
3. Example validation is carried out based on real power load data. And the proposed model is compared with other state-of-the-art models. The results show that the model has a good ability to fight against perturbations, and can accomplish the task of short-term power load prediction with stable prediction results.

The remaining sections of this paper are organized as follows. Section II introduces the time series decomposition model combining EWT and the time window approach. Section III presents the optimized Bi-LSTM algorithm based on FGSM and introduces a novel loss function. Section IV conducts simulation experiments using various datasets,

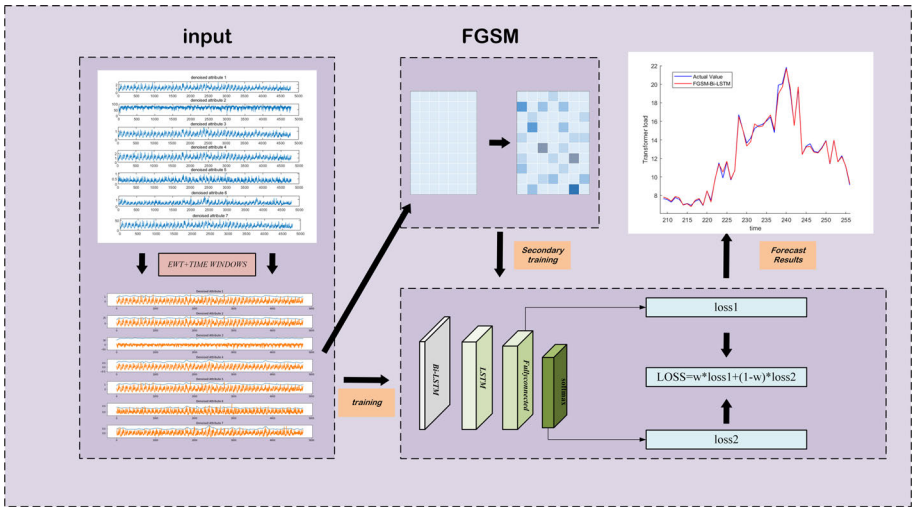


FIGURE 1. The designed algorithm framework.

compares the performance of the proposed model with other forecasting models, and validates its superiority. Section V concludes the paper.

II. TIME SERIES DECOMPOSITION ALGORITHM

Power load forecasting data is often affected by various interference and noise factors, resulting in instability and inaccuracy. Therefore, performing denoising on the data is necessary prior to forecasting, to ensure reliability and accuracy. Empirical Wavelet Transformation (EWT) is an automated signal processing method used for feature extraction and denoising of non-stationary and non-linear signals [32]. Unlike traditional wavelet analysis methods, EWT utilizes adaptive filters to decompose the signal progressively based on its local characteristics. Since EWT has limited degrees of freedom in wavelet selection, it adopts the principles of constructing Littlewood-Paley (Eq. 1) and Meyer (Eq. 2) wavelets to define the empirical scaling and wavelet functions:

$$\hat{\phi}_n(\omega)=\begin{cases}1 & \text{if}|\omega|\le(1-\gamma)\omega_n\\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n))] & \\ & \text{if}(1-\gamma)\omega_n\le|\omega|\le(1+\gamma)\omega_n\\ 0 & \text{else}\end{cases} \tag{1}$$

$$\hat{\psi}_n(\omega)=\begin{cases}1 & \text{if}(1+\gamma)\omega_n\le|\omega|\le(1-\gamma)\omega_{n+1}\\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_{n+1}))] & \\ & \text{if}(1-\gamma)\omega_{n+1}\le|\omega|\le(1+\gamma)\omega_{n+1}\\ \sin[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n))] & \\ & \text{if}(1-\gamma)\omega_n\le|\omega|\le(1+\gamma)\omega_n\\ 0 & \text{else}\end{cases} \tag{2}$$

where, the normalized frequency is denoted as ω, and it possesses periodicity. In order to satisfy the Shannon criteria, we make $\omega \in [0,\pi]$. γ is an important parameter to ensure that the overlap region between two consecutive state intervals is minimized, and its value is determined by the calculated boundary values as shown in Eq. (3):

$$\gamma < \min\left\{\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n}\right\}, \quad \gamma\in[0,1], \tag{3}$$

$$\beta(x)=x^4(35-84x+70x^2-20x^3). \tag{4}$$

The reconstructed original signal can be expressed as Eq. (5):

$$f(t)=W_f^t(0,t)^*\phi_1(t)+\sum_{n=1}^{N}W_f^t(n,t)^*\psi_n(t), \tag{5}$$

In Eq. (5), * is the convolution operation, $W_f^t(0,t)$ is the approximate coefficient after Fourier transform, and $W_f^t(n,t)$ is the detail coefficient after Fourier transform.

Due to the issue of data loss in prediction when applying signal decomposition algorithms directly to time series, in order to avoid the loss of future data in decomposition-based prediction models, this paper adopts a combination of moving window and EWT to denoise the data. Fig. 2 shows the noise reduction process flow and the specific steps are as follows:

1. Segment the time series data according to a certain window size, and each segment is a subsequence.
2. For each subsequence, perform the EWT transformation.
3. For each scale factor, calculate its local variance, which is used to determine whether the scale factor is noise or not.
4. Scale factors with local variance below the threshold are set to 0 according to a preset threshold value.

5. For each subsequence, perform the inverse EWT transformation to obtain the denoising subsequence.
6. Merge the denoised subsequences into a new sequence.

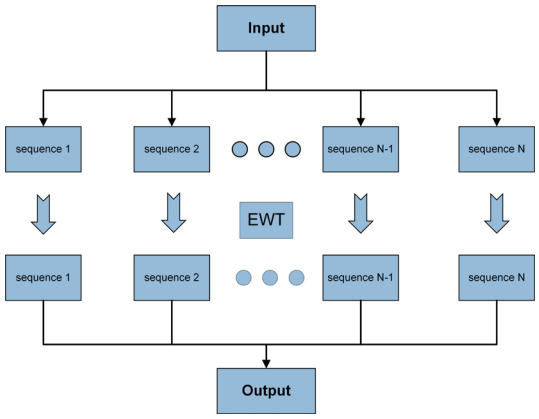


FIGURE 2. Noise reduction process flowchart.

III. MODEL ARCHITECTURE

A. LSTM NEURAL NETWORK

LSTM is a network model based on the improvement of the RNN (Recurrent Neural Network) algorithm and is widely used for processing time series data. To alleviate the issues of gradient vanishing and exploding that often occur in RNNs, LSTM introduces three gate units (input gate, output gate, forget gate) to control the flow of information transmission. This allows the network to selectively retain and forget information, thereby better capturing long-term dependencies in the sequence. Fig. 3 shows the structure of the traditional LSTM.

The structure of RNN represented by Eq. (6) to (7).

$$O_t = g(V \cdot S_t) \tag{6}$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \tag{7}$$

where, O_t represents the output at time t, and S_t represents the value of the hidden layer at time t. W is the weight matrix to represent the weights between each time point.

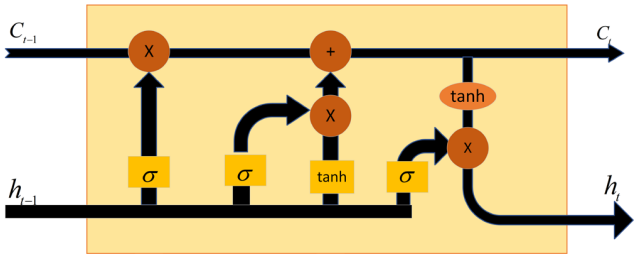


FIGURE 3. LSTM network structure.

During the training process of the LSTM neural network, the data features at time step t are initially input into the input layer, where they undergo processing by the activation function. Subsequently, at time step t, the data, the output of the hidden layer at time step *t-1*, and the information stored in the cell unit at time step *t-1* are fed into the nodes of the LSTM structure. These data are then processed by the Input Gate, the Output Gate, the Forget Gate, and the cell unit within the node before being passed on to the next node or the output layer. Finally, the backpropagation error is computed, and individual weights are updated. The calculation formulas are presented in Eqs. (8) to (13):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{8}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{9}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_C), \tag{10}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \tag{11}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \tag{12}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{13}$$

where, σ represents the sigmoid function. h_{t-1} is the output of the previous state. x_t is the input for the current state. W_f, W_i, W_c, W_o represent the weight coefficient matrices. h_t is the output of the current state, and $[h_{t-1}, x_t]$ indicates the concatenation of two matrices.

B. BI-LSTM NEURAL NETWORK

In recent years, Bidirectional Long Short-Term Memory (Bi-LSTM) networks have been widely adopted in domains such as natural language processing, speech recognition, and time series analysis. This is because Bi-LSTMs can learn temporal dependencies in both directions and have shown superior performance compared to unidirectional LSTM models. Different from traditional LSTM networks, Bi-LSTM networks can learn both forward and backward time dependencies by adding a reverse LSTM layer. This is very helpful for mining the long-term trend and short-term fluctuation characteristics in the grid load time series. The forward and inverse LSTMs capture different patterns of time series, and finally the outputs of both of them are combined together, which can make the prediction more accurate and robust. Fig. 4 show the structure of the Bi-LSTM.

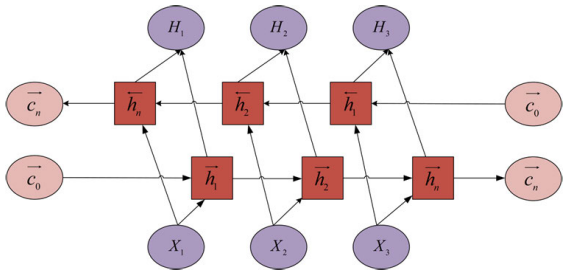


FIGURE 4. Bi-LSTM network structure.

Specifically, the structure of Bi-LSTM network can be divided into two parts: forward LSTM layer ($\overrightarrow{h_t}$) and reverse LSTM layer ($\overleftarrow{h_t}$). The forward LSTM layer handles the time step from the beginning to the end of the input sequence, while the reverse LSTM layer handles the time step from

the end to the beginning of the sequence. In this way, the Bi-LSTM network can capture both past and future contextual information to understand patterns and features in sequence data more comprehensively. These two parts can be represented as follows:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(h_{t-1}, x_t, c_{t-1}), \quad t \in [1, T] \tag{14}$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(h_{t-1}, x_t, c_{t-1}), \quad t \in [T, 1] \tag{15}$$

$$H_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{16}$$

C. FGSM GENERATES ADVERSATIAL SAMPLES

FGSM is widely used in the field of deep learning and machine learning as a simple but effective technique for generating adversarial samples. Its core idea is to deceive the deep learning model into generating false outputs by adding tiny perturbations to the input data. Its perturbation is shown in Eq. (17), which is a key parameter in adversarial attacks that determines the magnitude of the generated adversarial samples. Different perturbation sizes can generate adversarial attacks of different strengths, ranging from small changes to more significant perturbations.

$$\eta = \in sign(\nabla_x J(\theta, x, y)) \tag{17}$$

where $\in$ represents the adversarial strength, x denotes the input, y is the label of x, θ represents the model parameters, J is the loss function, and $\nabla_x J(\theta, x, y)$ represents the gradient of the loss function with respect to x. By linearizing the loss function around the current value of q. The optimal max-norm constrained perturbation of the perturbation η can be obtained.

In order to solve the problem of lack of labeling of data in the regression task, we classify the load status of transformers into three categories of normal, heavy load, and severe overload. This classification is based on the grid company's load evaluation standard for transformers. For the loss function, we utilize cross entropy, and its specific calculation rules are detailed below:

$$loss_1 = -\log(y|x + \eta; \theta), \tag{18}$$

$$y(\phi) = \begin{cases} 1, & 0 < \phi \leq 10 \\ 2, & 10 < \phi \leq 20 \\ 3, & \phi > 20, \end{cases} \tag{19}$$

where, $loss_1$ is the cross entropy. y represents different load states: 1 indicates to a normal load state, 2 indicates a heavy load state, and 3 indicates a severe overload state. ϕ is the transformer load factor.

D. BI-LSTM ALGORITHM BASED ON FGSM

In this paper, we propose a hybrid model, referred to as the FGSM-Bi-LSTM hybrid model, which takes time series data as its input. To ensure the reliability and accuracy of the data, this paper refers to the word vector representation method used in Natural Language Processing (NLP). We employ a sliding window model to concatenate the load data at specific time points with their relevant features, creating a vector representation.

Additionally, the EWT method is applied to perform denoising on the data. Madry et al., have suggested that in networks with high linearity such as LSTM, the presence of perturbed data can have a significant impact on the results [29].

To enhance the network's defense capability, we incorporate the FGSM to generate perturbations and attack the model. To prevent overfitting of the training data, an Early Stopping algorithm is introduced to monitor the performance of the model on the validation set. The training process is promptly terminated when no further improvement is observed on the validation set. The Early Stopping algorithm pseudo-code is as follows:

Algorithm 1 Early Stopping

```
best_loss ← ∞
patience ← 0
max_patience ← 10
while ( epoch < num_epochs ) do
    train_model ()
    validation_loss ← calculate_loss(validation_data)
    if ( validation_loss < best_loss ) then
        best_loss ← validation_loss
        patience ← 0
    else
        patience ← patience + 1
    end if
    if (patience ≥ max_patience) then
        break
    end if
end while
```

E. MODEL EVALUATION METRICS

To improve the accuracy of the prediction, we first normalize the individual feature sequences as follows:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{20}$$

where x_i is the original data, x_i' is the preprocessed data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum values respectively. The predicted values are obtained and then reverse-normalized to obtain the predicted load values.

Furthermore, to assess the model's validity, we employ a set of evaluation metrics. Among them, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Pearson correlation coefficient (r) are commonly used in regression tasks to evaluate the goodness of the model. MAE is a statistical metric for assessing the performance of a predictive model. Smaller MAE values indicate higher accuracy and closer proximity of the model's predictions to the actual observations. In contrast to MAE, which squares errors and thus magnifies the impact of outliers in MSE, MAE assesses the magnitude of the average

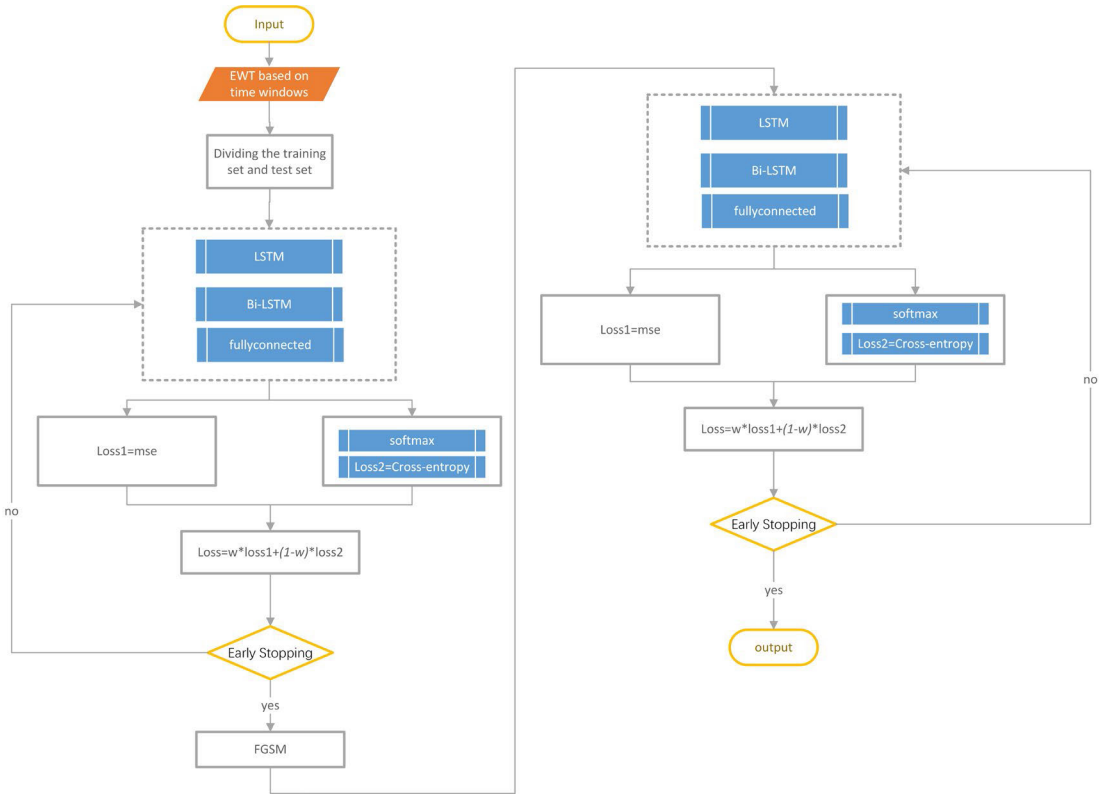


FIGURE 5. FGSM-Bi-LSTM flowchart.

prediction error of the model. Conversely, RMSE involves taking the square root of MSE, effectively eliminating the squared error. Meanwhile, the Pearson correlation coefficient (r) is employed to quantify the strength and direction of the relationship between two variables. The specific rules for calculating the above metrics are shown below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{21}$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}, \tag{22}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \tag{23}$$

$$r(y_i, \hat{y}_i) = \frac{\mathrm{cov}(y_i, \hat{y}_i)}{\sqrt{D(y_i)}\sqrt{D(\hat{y}_i)}}, \tag{24}$$

$$loss_2 = MSE(predicted, true), \tag{25}$$

$$LOSS = (1 - w) \cdot loss_1 + w \cdot loss_2. \tag{26}$$

where n represents the number of samples, y_i is the true value at time point i, and $\hat{y}_i$ is the predicted value at time point i. Furthermore, FGSM is employed to generate adversarial samples. Subsequently, a new loss function is constructed by combining the regression loss and the cross-entropy weighted sum, as depicted in Eqs. (25) and (26), and where w represents the weigh. Finally, the flowchart of the proposed FGSM-Bi-LSTM model is depicted in Fig. 5.

IV. SIMULATION EXPERIMENTS AND ANALYSIS

In this section we use real transformer data from a provincial capital city in Southwest China for the case study. Firstly, the power load data used is introduced, secondly, the noise reduction process and the setting of the weights of the loss function are described, and finally, the designed FGSM-Bi-LSTM model is compared with the ARIMA [5], QR-GRU network [7], LSTM network [11], the Transformer [15], XGBoost [33] models are compared cross-sectionally to study the superiority of the proposed model. The key parameters in the proposed FGSM-Bi-LSTM model are shown in Table 1.

A. DATA DESCRIPCTIONS

Transformer data from three locations in Southwest China are used to evaluate the accuracy of various input feature set scenarios and the algorithms proposed in this paper. The three transformers are uniquely identified by the following code numbers: "47691280," "54649268," and "72541267," with their respective utilization numbers presented in Table 2. Each transformer dataset comprises a total of 4,780 real-time monitoring data points recorded between February 1, 2023, and March 22, 2023, with a 15-minute sampling interval. The selected parameters for prediction include three-phase

TABLE 1. Parameters of FGSM-Bi-LSTM.

Parameters	Value
Epsilon	0.1
Epoch	100
Mini Batch Size	16
Max Patience	20
Test Interval	5
LSTM's hidden Size	40
Bi-LSTM's hidden Size	60

maximum current, maximum transformer loading rate, total unbalance rate, current three-phase average, phase A current, phase B current, and phase C current, all aimed at predicting real-time transformer load. In addition, other parameters of the transformer used in the experiment are shown in Table 3.

TABLE 2. Transformers' number.

Transformer No.	unique identification code
T-I	47691280
T-II	54649268
T-III	72541267

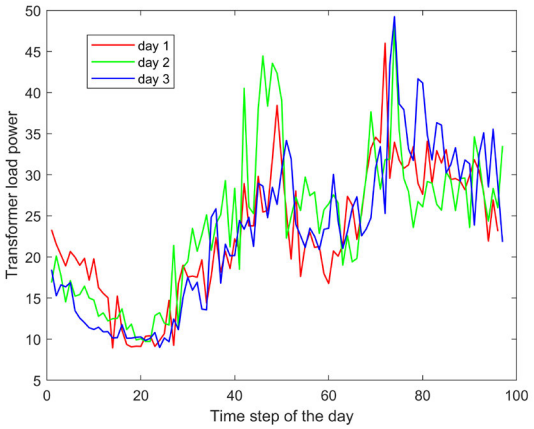
TABLE 3. Transformer parameters.

Parameters	Value
3 phases (HZ)	50
Rated voltage (kV)	550/18
Rated capacity (kVA)	723000
Short circuit resistance (%)	16.24
No-load current (%)	0.04
No-load loss (kW)	244.4
Load loss (kW)	1151

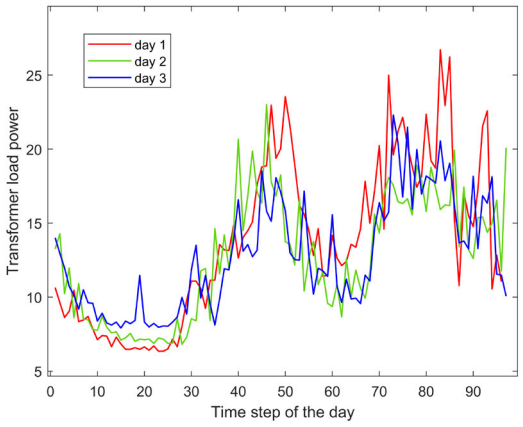
Further, in order to visualize the complex dynamics of the power load data, we visualized the load data for the three transformers used as shown in Fig. 6. It can be seen that the trends and characteristics of the electrical loads on any given day are very close to each other, which indicates a strong periodicity and regularity in the changes of electrical loads over a 24-hour period. In addition, the more complex dynamics also make the electrical load more difficult to realize. For this reason, we describe in the following the processing operations that we introduce in our forecasting model.

B. SLIDING WINDOW-BASEDEWT DENOISING

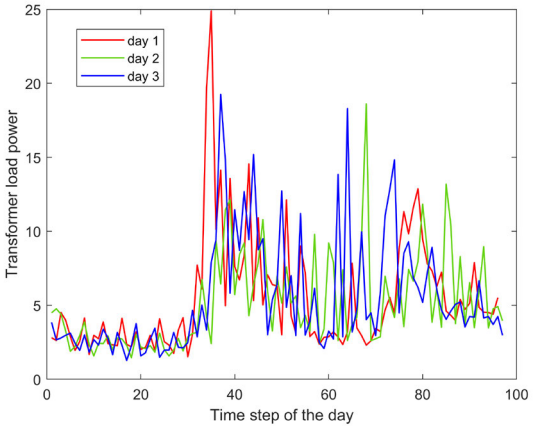
First, we introduce a time window based EWT noise reduction method. Due to the potential data loss caused by EWT during data processing, this study combines the sliding window algorithm with EWT to denoise the data. The denoising ability of EWT increases with the number of decomposition



(a) Electrical loads of transformer I



(b) Electrical loads of transformer II



(c) Electrical loads of transformer III

FIGURE 6. Electrical loads of three transformers.

levels, but it may also result in some loss of signal information. The decomposition level in this study is N = 2, the size of the sliding window is 30, and the overlap rate is 0.1. To ensure that the length of the data does not change, the step size is shown in Eq. (27) and (28):

$$o = \lceil z \cdot l \rceil , \tag{27}$$

$$s = \lceil z - o \rceil , \tag{28}$$

where o is the overlap length, s is the step size, and [ ] denotes the ceiling function. The processed attributes are shown in Fig. 7.

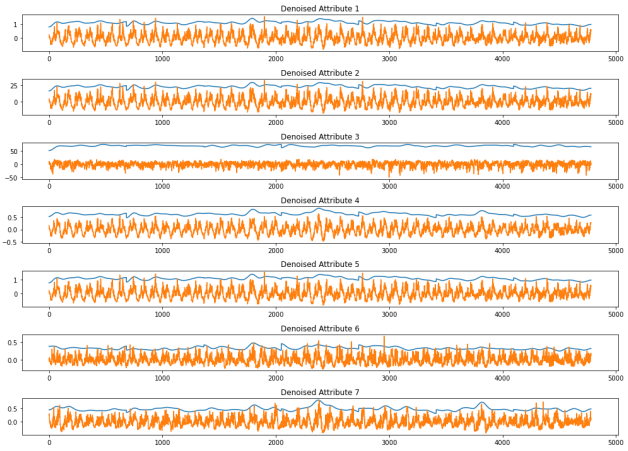


FIGURE 7. Visualization of each property after processing.

C. LOSS FUNCTION WEIGHT SETTING

To improve prediction accuracy, this study adjusts the weights in the loss function (Eq. 26) and observes changes in the four evaluations metrics from Section III to determine the optimal values for the weights. The weight w is sampled with a step size of 0.1 within the range of 0 to 1. For each weight, the FGSM-Bi-LSTM model is trained on a subset of 208 data samples from the dataset of *T-I*, with 48 data samples used as a validation set. The model is optimized for 50 iterations, and the average values of the evaluation metrics are obtained. The changes in the evaluation metrics for different weights are shown in Fig. 8.

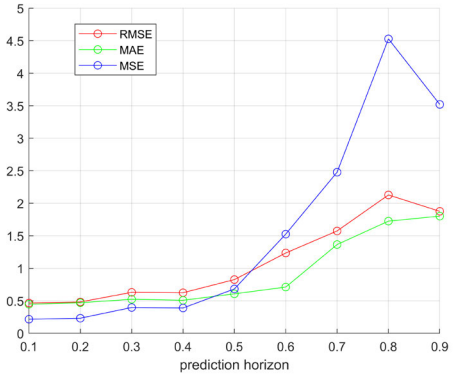


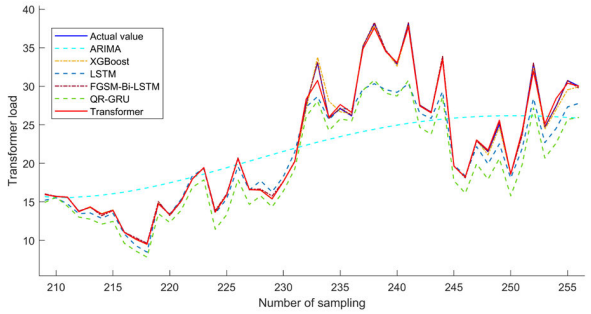
FIGURE 8. Evaluation indicators under different w.

As shown in Fig. 8, the RMSE and MAE of the predicted results generally exhibit an upward trend as the weight w for MSE increases. This is because a smaller adversarial loss indicates higher accuracy on the original data. To enhance robustness against adversarial examples, a higher weight should be assigned to the adversarial loss. And to achieve a balance between robustness and accuracy, w is set to 0.1 in the experiments. This choice of weight aims to ensure that the model maintains good robustness while still achieving satisfactory accuracy.

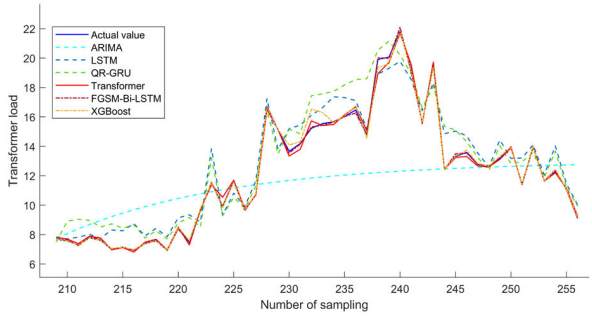
D. COMPARISON OF PREDICTION RESULTS

In this subsection, we compare the FGSM-BI-LSTM network with several other models commonly used for power load forecasting. Since excessively long data sequences could hinder the visualization of algorithmic performance differences, we randomly select a sequence consisting of 208 data points from the dataset as input and use 48 data points as output. This subset of data provides a more manageable and representative sample for comparing the performance of different models. For reliable experimental results, each model was run 30 times independently and averaged to obtain the performance metrics for each algorithm.

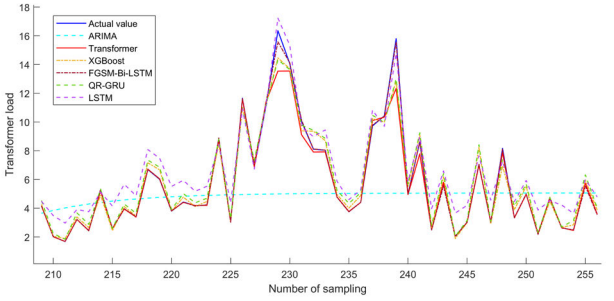
The results are shown in Table 4. Fig. 9 and Fig. 10 show the comparison of the prediction performance of the six models in three transformers dataset.



(a) Load forecast curves of six models for *T-I*



(b) Load forecast curves of six models for *T-II*



(d) Load forecast curves of six models for *T-III*

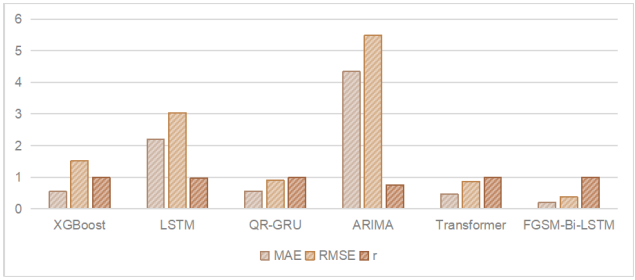
FIGURE 9. Electricity load forecasting curves.

The difference in performance of the different models can be found in Fig. 9. The ARIMA model is more poorly predicted and is only able to follow the general trend of the net load series, which makes it difficult to adapt to the dramatic fluctuations of the series. The overall trend of LSTM prediction is consistent with the actual data, but there is a certain delay, resulting in a decrease in prediction accuracy. This may be due to the fact that LSTM processes time series sequentially, captures historical information gradually, and is also affected by data noise. The QR-GRU model performs poorly on T-II and T-III predictions, mainly due to the strong randomness of the dataset, while the model performs poorly in capturing short-term trends and resisting noise. In contrast, the Transformer and XGBoost models performed relatively well on all three transformers, but neither of them predicted accurately enough at the moment when the data fluctuated sharply. Both models perform well when dealing with long sequences, but there may be problems with overfitting or difficulty capturing valid information on short sequences. In contrast, the prediction of FGSM-Bi-LSTM is the closest to the actual value and can effectively track the actual load trend. This is due to the fact that the FGSM module enhances the noise resistance of the model, the EWT decomposition reduces the data noise, and the Bi-LSTM captures both past and future information, which improves the prediction accuracy.

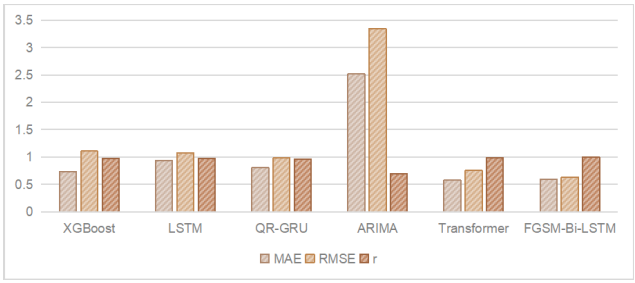
TABLE 4. Performance indicators for each algorithm.

No.	Model	*MAE*	*RMSE*	*MSE*	*r*
T-I	XGBoost	0.544	1.515	2.295	0.983
	LSTM	2.203	3.040	9.241	0.975
	QR-GRU	0.563	0.897	0.805	0.986
	ARIMA	4.340	5.482	30.052	0.758
	Transformer	0.458	0.863	0.744	0.993
	FGSM-Bi-LSTM	**0.206**	**0.374**	**0.140**	**0.999**
T-II	XGBoost	0.730	1.106	1.223	0.976
	LSTM	0.939	1.076	1.157	0.975
	QR-GRU	0.803	0.981	0.963	0.956
	ARIMA	2.515	3.348	11.209	0.697
	Transformer	0.577	0.757	0.573	0.984
	FGSM-Bi-LSTM	**0.589**	**0.633**	**0.401**	**0.999**
T-III	XGBoost	0.863	1.214	1.474	0.978
	LSTM	0.746	1.433	2.053	0.982
	QR-GRU	1.032	1.530	2.340	0.924
	ARIMA	2.592	3.633	13.199	0.318
	Transformer	0.655	0.865	0.748	0.990
	FGSM-Bi-LSTM	**0.493**	**0.423**	**0.179**	**0.991**

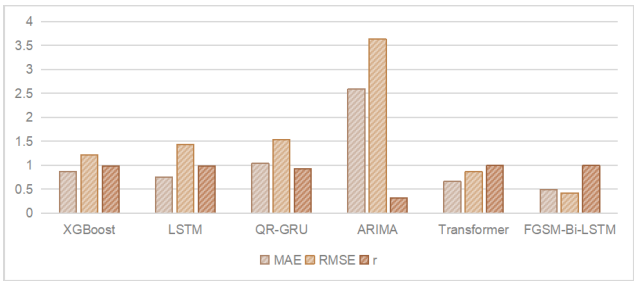
Analyzing the specific experimental results in Table 4 and Fig. 10, it is evident that the FGSM-Bi-LSTM model, proposed in this study, outperforms the Transformer, LSTM, QR-GRU, ARIMA, and XGBoost models across various performance metrics. The proposed FGSM-Bi-LSTM model reduces MAE by 41.44%, 53.95%, 47.43%, 84.27% and 26.55% compared to Transformer, LSTM, QR-GRU, ARIMA and XGBoost models, respectively. Similarly, it reduces RMSE by 61.08%, 66.45%, 55.38%, 87.54% and 42.54%, and MSE by 85.58%, 94.22%, 82.47%, 98.68% and 65.13%. The Pearson correlation coefficient r is improved by 0.0173, 0.0190, 0.0410, 0.4053 and 0.0073.



(a) Prediction performance of six models for *T-I*



(b) Prediction performance of six models for *T-II*



(c) Prediction performance of six models for *T-III*

FIGURE 10. Predictive performance of five models in three regions.

The designed FGSM-Bi-LSTM model excels in performance across all evaluation metrics. The ARIMA model, although capable of predicting the general trend of the data, struggles with short-term transformer loading data characterized by nonlinearity, leading to less accurate predictions. This limitation becomes more pronounced when dealing with drastic fluctuations in the data. In short-term transformer load prediction, the load tends to fluctuate with the time of day, introducing noise that hampers model accuracy. On the other hand, XGBoost, Transformer, and QR-GRU algorithms manage to capture data changes to some extent, but they struggle to accurately predict peaks. Among deep learning algorithms, LSTM performs the poorest, despite its general trend prediction being relatively close to the actual sequence. However, its MSE, RMSE, and MAE values are larger.

In summary, the model proposed in this study employs signal processing techniques to reduce noise in the original data. FGSM improves model robustness against adversarial samples, while Bi-LSTM captures time series neighborhood information, enhancing peak and overall accuracy.

E. ANALYSIS OF MODEL STABILITY

In this subsection, we analyze the stability of the proposed FGSM-Bi-LSTM model. The stability of the model is judged by setting the input step size of the model to 208 and by changing the prediction step size to 48 (12h), 96 (24h), 144 (36h), 192 (48h), 240 (60h). The experimental results are shown in Table 5 and the visualization results are shown in Fig. 11.

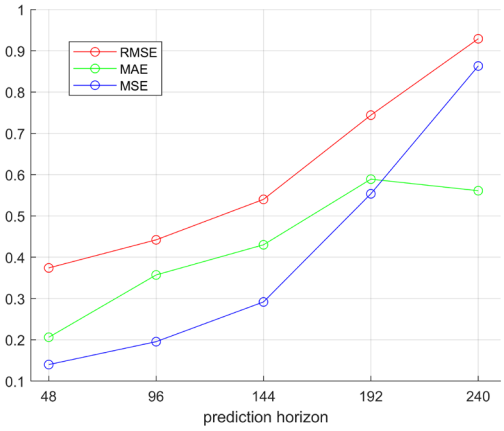
As shown in Table 5, RMSE, MAE, and MSE gradually increase with longer prediction steps. However, the correlation coefficient r remains near 0.998, indicating strong agreement between predictions and actual data. Even when the prediction length exceeds the input length, RMSE and MAE remain low at 0.929, 0.974, 0.767 and 0.561, 0.771, 0.546 for the three datasets respectively. In summary, the results show that the designed model is still valid and accurate in making longer predictions.

From Fig. 11, it can be observed that with the increase of the prediction length, the change trend of the three evaluation indexes is smooth. This indicates that the proposed model has good stability and is capable of predicting different ranges of short-term loads. When the prediction length is increased from 48 to 240, the model error is maintained in a small range, which indicates that the model can effectively extract the trend and periodicity features in the data. The EWT noise reduction for data preprocessing also helps the model to capture the information better. Meanwhile, the FGSM adversarial training enhances the anti-interference ability of the model and improves the prediction accuracy.

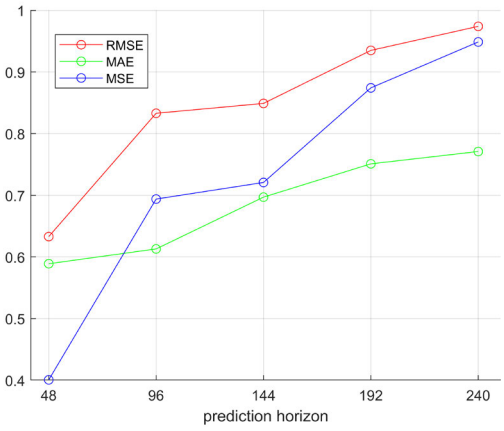
TABLE 5. Comparison of results for different predicted lengths.

No.	Index	48	96	144	192	240
T-I	RMSE	0.374	0.442	0.540	0.744	0.929
	MAE	0.206	0.357	0.430	0.589	0.561
	MSE	0.140	0.195	0.292	0.554	0.863
	r	0.999	0.999	0.999	0.999	0.999
T-II	RMSE	0.633	0.833	0.849	0.935	0.973
	MAE	0.589	0.613	0.697	0.751	0.771
	MSE	0.401	0.693	0.722	0.875	0.948
	r	0.999	0.999	0.999	0.999	0.999
T-III	RMSE	0.423	0.547	0.645	0.725	0.766
	MAE	0.493	0.327	0.435	0.498	0.545
	MSE	0.179	0.299	0.416	0.5259	0.5875
	r	0.999	0.999	0.999	0.999	0.999

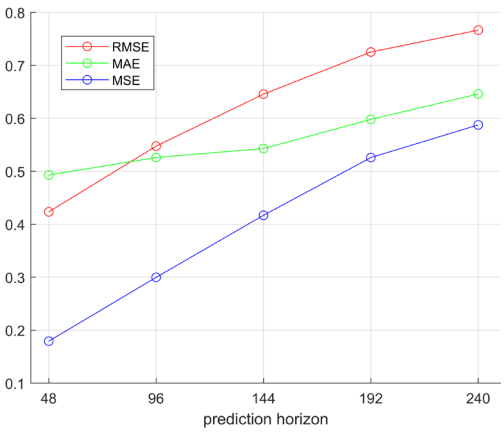
In summary, the joint preprocessing of EWT and time window is used to effectively reduce the data noise and provide clearer input for the model. The deep network based on Bi-LSTM improves the extraction of data features and enhances the predictive stability of the model by capturing the contextual information in the time dimension. In addition, the FGSM module enhances the ability of the model to resist perturbations and makes it more robust. Finally, Early Stopping algorithm is applied to avoid model overfitting and improve the generalization ability of the model to new data. The methods can work synergistically to enhance model inputs, architectures, training, and generalization for superior forecasting capabilities compared to existing models.



(a) Prediction accuracy of *T-I* at various lengths of time



(b) Prediction accuracy of *T-II* at various lengths of time



(c) Prediction accuracy of *T-III* at various lengths of time

FIGURE 11. Prediction accuracy at different time lengths.

V. CONCLUSION

This study discusses the short-term load forecasting problem in smart grids and proposes a short-term transformer load forecasting model based on Bi-LSTM. We use EWT to decompose the time series data of transformer loads and train the model using a deep neural network based on Bi-LSTM. We also introduce the FGSM algorithm for data augmentation. Finally, we combine cross-entropy and

mean squared error and determine their respective weights to construct a new loss function for retraining the model. In order to verify the superiority and robustness of the proposed model, we compare the FGSM-Bi-LSTM model proposed in this paper with five models, namely, LSTM, QR-GRU, ARIMA, XGBoost, and Transformer, on three datasets. Example experimental results show that the FGSM-Bi-LSTM model proposed in this paper has the following advantages:

1. By combining the signal processing algorithm EWT with time windows, the algorithm effectively reduces noise in the data and significantly improves the prediction accuracy.

2. The FGSM algorithm and the retraining step enhance the model's defense against perturbations.

3. The deep neural network based on Bi-LSTM can capture local features and trends in the data, resulting in high prediction accuracy.

The proposed algorithm exhibits good stability and high prediction accuracy, making it suitable for short-term and ultra-short-term load forecasting tasks for transformers. In future research, we will consider the relationships between features, as well as the relationships between features and targets, to improve the network structure. We will also compare the proposed model with other advanced techniques.

REFERENCES

[1] A. Heydari, M. M. Nezhad, E. Pirshayan, D. A. Garcia, F. Keynia, and L. De Santoli, "Short-term electricity price and load forecasting in isolated power grids based on composite neural network and gravitational search optimization algorithm," *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115503, doi: 10.1016/j.apenergy.2020.115503.

[2] N. Jha, D. Prashar, M. Rashid, S. K. Gupta, and R. K. Saket, "Electricity load forecasting and feature extraction in smart grid using neural networks," *Comput. Electr. Eng.*, vol. 96, Dec. 2021, Art. no. 107479, doi: 10.1016/j.compeleceng.2021.107479.

[3] K. Ni, J. Wang, G. Tang, and D. Wei, "Research and application of a novel hybrid model based on a deep neural network for electricity load forecasting: A case study in Australia," *Energies*, vol. 12, no. 13, p. 2467, Jun. 2019, doi: 10.3390/en12132467.

[4] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustain. Energy, Grids Netw.*, vol. 6, pp. 91–99, Jun. 2016, doi: 10.1016/j.segan.2016.02.005.

[5] E. Erdogdu, "Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey," *Energy Policy*, vol. 35, no. 2, pp. 1129–1146, Feb. 2007, doi: 10.1016/j.enpol.2006.02.013.

[6] H. Liu, "The forecast of household power load based on genetic algorithm optimizing BP neural network," *J. Phys., Conf. Ser.*, vol. 1871, no. 1, Apr. 2021, Art. no. 012110, doi: 10.1088/1742-6596/1871/1/012110.

[7] C. Sun, Q. Lv, S. Zhu, W. Zheng, Y. Cao, and J. Wang, "Ultra-short-term power load forecasting based on two-layer XGBoost algorithm considering the influence of multiple features," *High Voltage Eng.*, vol. 47, pp. 2885–2898, Aug. 2021.

[8] Y. Chen and H. Tan, "Short-term prediction of electric demand in building sector via hybrid support vector regression," *Appl. Energy*, vol. 204, pp. 1363–1374, Oct. 2017, doi: 10.1016/j.apenergy.2017.03.070.

[9] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018, doi: 10.1109/TSG.2017.2686012.

[10] Z. Niu, Z. Yu, W. Tang, Q. Wu, and M. Reformat, "Wind power forecasting using attention-based gated recurrent unit network," *Energy*, vol. 196, Apr. 2020, Art. no. 117081, doi: 10.1016/j.energy.2020.117081.

[11] D. Ageng, C.-Y. Huang, and R.-G. Cheng, "A short-term household load forecasting framework using LSTM and data preparation," *IEEE Access*, vol. 9, pp. 167911–167919, 2021, doi: 10.1109/ACCESS.2021.3133702.

[12] S. Atef, K. Nakata, and A. B. Eltawil, "A deep bi-directional long-short term memory neural network-based methodology to enhance short-term electricity load forecasting for residential applications," *Comput. Ind. Eng.*, vol. 170, Aug. 2022, Art. no. 108364, doi: 10.1016/j.cie.2022.108364.

[13] K. Li, W. Huang, G. Hu, and J. Li, "Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network," *Energy Buildings*, vol. 279, Jan. 2023, Art. no. 112666, doi: 10.1016/j.enbuild.2022.112666.

[14] N. Jha, P. Deepak, M. Rashid, S. K. Gupta, and R. K. Saket, "Electricity load forecasting and feature extraction in smart grid using neural networks," *Comput. Electr. Eng.*, vol. 99, Dec. 2021, Art. no. 107479, doi: 10.1016/j.compeleceng.2021.107479.

[15] Q. Zhang, J. Chen, G. Xiao, S. He, and K. Deng, "TransformGraph: A novel short-term electricity net load forecasting model," *Energy Rep.*, vol. 9, pp. 2705–2717, Dec. 2023, doi: 10.1016/j.egyr.2023.01.050.

[16] B. Vidakovic and C. B. Lozoya, "On time-dependent wavelet denoising," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2549–2554, Sep. 1998, doi: 10.1109/78.709544.

[17] M. Last, Y. Klein, and A. Kandel, "Knowledge discovery in time series databases," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 31, no. 1, pp. 160–169, Feb. 2001, doi: 10.1109/3477.907576.

[18] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22–29, May 2016, doi: 10.1016/j.apenergy.2016.02.114.

[19] A. Laouafi, M. Mordjaoui, F. Laouafi, and T. E. Boukelia, "Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology," *Int. J. Electr. Power Energy Syst.*, vol. 77, pp. 136–144, May 2016, doi: 10.1016/j.ijepes.2015.11.046.

[20] Y. Luo, Q. Cheng, S. Yan, and D. Yang, "Situation awareness method of the distribution network based on EMD-SVD and Elman neural network," *Energy Rep.*, vol. 8, pp. 632–639, Nov. 2022, doi: 10.1016/j.egyr.2022.05.212.

[21] D. Pan, B. Xu, J. Ma, Q. Ding, J. Ding, J. Zhang, and Q. Zhang, "Short-term load forecasting based on EEMD-approximate entropy and ELM," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*. Beijing, China: IEEE, Nov. 2019, pp. 1772–1775, doi: 10.1109/iSPEC48194.2019.8974925.

[22] D. Deng, J. Li, Z. Zhang, Y. Teng, and Q. Huang, "Short-term electric load forecasting based on EEMD-GRU-MLR," *Power Syst. Technol.*, vol. 44, no. 2, pp. 593–602, 2020, doi: 10.13335/j.1000-3673.pst.2019.0113.

[23] I. Niazazari and H. Livani, "Attack on grid event cause analysis: An adversarial machine learning approach," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*., Washington, DC, USA: IEEE, Feb. 2020, pp. 1–5, doi: 10.1109/ISGT45199.2020.9087649.

[24] Y. Chen, Y. Tan, and D. Deka, "Is machine learning in power systems vulnerable?" in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*. Aalborg, Denmark: IEEE, Oct. 2018, pp. 1–6, doi: 10.1109/SmartGridComm.2018.8587547.

[25] J. Li, Y. Liu, T. Chen, Z. Xiao, Z. Li, and J. Wang, "Adversarial attacks and defenses on cyber–physical systems: A survey," *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5103–5115, Jun. 2020, doi: 10.1109/JIOT.2020.2975654.

[26] A. Sayghe, J. Zhao, and C. Konstantinou, "Evasion attacks with adversarial deep learning against power system state estimation," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*. Montreal, QC, Canada: IEEE, Aug. 2020, pp. 1–5, doi: 10.1109/PESGM41954.2020.9281719.

[27] J. Tian, B. Wang, Z. Wang, K. Cao, J. Li, and M. Ozay, "Joint adversarial example and false data injection attacks for state estimation in power systems," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13699–13713, Dec. 2022, doi: 10.1109/TCYB.2021.3125345.

[28] Y. Chen, Y. Tan, and B. Zhang, "Exploiting vulnerabilities of load forecasting through adversarial attacks," in *Proc. 10th ACM Int. Conf. Future Energy Syst.* Phoenix AZ USA: ACM, Jun. 2019, pp. 1–11, doi: 10.1145/3307772.3328314.

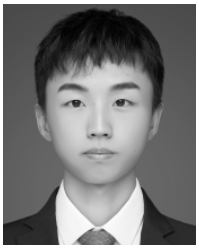
[29] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," Sep. 2017, *arXiv:1706.06083*. Accessed: Jul. 10, 2023.

[30] M. Gallagher, N. Pitropakis, C. Chrysoulas, P. Papadopoulos, A. Mylonas, and S. Katsikas, "Investigating machine learning attacks on financial time series models," *Comput. Secur.*, vol. 123, Dec. 2022, Art. no. 102933, doi: 10.1016/j.cose.2022.102933.

[31] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Müller, "Adversarial attacks on deep neural networks for time series classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Budapest, Hungary, 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8851936.

[32] R. Gao, L. Du, P. N. Suganthan, Q. Zhou, and K. F. Yuen, "Random vector functional link neural network based ensemble deep learning for short-term load forecasting," *Expert Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117784, doi: 10.1016/j.eswa.2022.117784.
[33] Z. Yu, Y. Sun, J. Zhang, Y. Zhang, and Z. Liu, "Gated recurrent unit neural network (GRU) based on quantile regression (QR) predicts reservoir parameters through well logging data," *Frontiers Earth Sci.*, vol. 11, Jan. 2023, Art. no. 1087385, doi: 10.3389/feart.2023.1087385.

QINGCHAN LIU was born in Cuiping, Sichuan, in August 1982. He received the degree in electrical engineering from North China Electric Power University. He is currently a Technical Expert and a Senior Engineer with the Measurement Center, Yunnan Power Grid Company Ltd. His research interests include electricity measurement and intelligent energy measurement technology, and online monitoring of measuring equipment operation status.



JIANING CAO was born in Taiyuan, Shanxi, China. He is currently pursuing the bachelor's degree in mechanical engineering with the Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming, China. His main research interests include operations research, decision-making, and machine learning.

JINGCHENG ZHANG was born in Qujing, Yunnan, China. He is currently pursuing the bachelor's degree in data science and big data technology with the Faculty of Science, Kunming University of Science and Technology, Kunming, China. His main research interests include swarm intelligence, data mining, and in-depth learning.

YAO ZHONG received the master's degree in test and measurement technology and instrumentation from the Wuhan University of Technology. He is currently a Senior Engineer with the Metering Center, Yunnan Power Grid Company Ltd. His research interests include power measurement and intelligent operation and maintenance.

TINGJIE BA received the master's degree from the Kunming University of Science and Technology, in July 2016. Since 2019, he has been the Senior Manager with the Marketing Department, Yunnan Power Grid Company Ltd. His main research interests include demand side response of new energy access, and optimization of electric interaction between vehicle and pile networks.

YIMING ZHANG received the degree in electrical engineering and automation from the Kunming University of Science and Technology, in 2014. He is currently an Engineer with the Metering Center (Power Load Control Technology Center), Yunnan Power Grid Company Ltd. His main research interests include power grid digitalization, power metering automation, and power monitoring system network security.