Explainable Spectrum Prediction Based on VMD-LSTM

Wei XU, Jianzhao ZHANG, Zhe SU

Sixty-third Research Institute, National University of Defense Technology,
No. 18, Houbiaoying, Nanjing, 210007, Jiangsu, China

yerenxw@163.com, jianzhao63s@nudt.edu.cn, zhes@nuaa.edn.cn

Submitted June 24, 2025 / Accepted November 17, 2025 / Online first December 22, 2025

Abstract. *To improve the accuracy and interpretability of neural network enabled spectrum prediction, an explainable spectrum prediction framework based on Variational Mode Decomposition (VMD) and Long Short-Term Memory (LSTM) networks, integrated with the Shapley Additive Explanations (SHAP) method (VMD-LSTM) is proposed in this work. Firstly, the raw spectrum data is decomposed into multiple Intrinsic Mode Functions (IMFs) via VMD to reduce sequence complexity. These IMFs are then fed into the LSTM network in parallel to improve prediction accuracy. Secondly, the SHAP method is incorporated to evaluate the impact weights of individual IMF components on the prediction outcomes, revealing the model's decision-making logic. Finally, we weight the input data by multiplying each IMF by its SHAP value to optimize prediction performance. Simulation results based on real spectrum data demonstrate that the proposed VMD-LSTM significantly outperforms baseline models on the metrics of Weighted Quality Evaluation Index (WQE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), mean absolute error (MAE). By incorporating SHAP weights to refine the model input features, the framework not only provides transparent explanations for the 'black-box' model but also reduces the average WQE, RMSE and MAPE by 3.99%, 3.23% and 3.67%, respectively.*

Keywords

Spectrum prediction, Shapley additive explanations, variational mode decomposition, explainable artificial intelligence

1. Introduction

The advent of 6G networks will drive unprecedented data generation across applications spanning mobile devices, sensors, and IoT systems [1]. 6G networks require high-speed, low-latency, and large-scale connectivity, thereby intensifying the demand for efficient wireless spectrum allocation. However, the limited spectrum resources desire efficient allocation and utilization. As an opportunistic spectrum technology, cognitive radio can greatly improve the spectrum utilization efficiency, thus alleviating the current scarcity of spectrum resources [2]. Nevertheless, the exponential growth in user numbers significantly increases sensing time and energy consumption, which potentially leads to spectrum underutilization. In order to solve the above problems, spectrum prediction has attracted more and more attention. By analyzing historical data to identify channel occupancy patterns and the potential correlation between channels, reducing energy costs and boosting throughput [2].

As a typical time series prediction problem, spectrum prediction methods can be divided into traditional machine learning methods and deep learning methods. Traditional approaches include regression model, Hidden Markov Model (HMM), Support Vector Machine (SVM), Random Forest (RF), and Kalman Filter (KF). These methods require a lot of prior knowledge and have limited performance in complex environment [3]. Deep learning models adaptively adjust complexity through parameter tuning, efficiently learning complex spectral patterns without prior knowledge. Previous studies have proved that neural network model is superior to classical prediction models [4].

The research on deep learning methods for time series prediction is mainly based on Recurrent Neural Network (RNN) and its improved model, which is the most effective scheme to mine the time correlation of spectral data [5].The Long Short-Term Memory (LSTM) network was adopted in [6] to predict the channel occupation in multiple time slots, which was shown to perform better than neural network and HMM models. Reference [7] proposed a prediction model based on time series data using CNN-Transformer and LSTM-Transformer architectures to collect time series data of patients 4, 8, and 12 hours before the diagnosis of sepsis for analysis. Reference [8] used Convolutional Neural Network (CNN) to mine local and regional spectral correlation, and adopted Gated Recurrent Unit (GRU) to find the time correlation in the input data. We adopted Wavelet Decomposition (WD) in [9] to partition spectral data into distinct frequency bands to improve the prediction accuracy. However, there exists the problem of spectrum aliasing among the sub-signals of WD. In this study, VMD effectively addresses the spectrum aliasing issue in WD subsignals, thereby enhancing both the prediction performance and interpretabil-

DOI: 10.13164/re.2026.0015

ity. Reference [10] combined VMD with LSTM and GRU, and realized accurate prediction of wave height by separating nonlinearity and non-smoothness in data through VMD. Furthermore, VMD-LSTM model has also been widely used in solar radiation flux [11], power load prediction [12], network flow estimation [13], wave height prediction [14] and other fields, but there are few relevant studies in the field of spectrum prediction.

While deep learning models outperform classical methods, their accuracy improvements often come at the cost of complexity, resulting in opaque "black-box" predictions and a lack of interpretability. However, understanding the reasons for the predictions is important for assessing trust in the model [15]. Explainable Artificial Intelligence (XAI), as the main method to solve this defect, has received extensive attention and research. XAI provides transparent explanations of model behavior, enabling users to understand decision-making processes [16]. Reference [17] used SHAP to help users interpret the predictions of the complex XGBoost model and examine the importance of various features affecting the prediction of gold price. The results showed that the prediction performance can be significantly improved by using the SHAP method. Reference [18] introduced a causal interpretable GRU, using SHAP values to quantify feature impacts for real-time predictions. The method was applied to actual engineering, and the results showed that this method had higher reliability and efficiency. Reference [19] used LSTM to predict the energy consumption of enterprises and introduced the XAI method to find out the parameters that have strong significance to the model output, which can provide support for energy allocation decisions.

The application of XAI methods in various scenarios has been well studied, but there are few related studies in the field of spectrum prediction. In this study, the SHAP method is employed to explain the LSTM prediction model based on VMD . It reveals the impact of IMFs on the prediction outcomes and optimizes the model performance by leveraging the calculated SHAP values. The main contributions of the study are as follows:

- Firstly, by incorporating VMD, the complexity of spectrum sequence and the influence of noise are reduced. The decomposed sub-sequences are imported in parallel into the LSTM model, achieving a significant improvement in prediction accuracy.
- Secondly, SHAP method is used to quantify the influence of each IMF on the prediction results and explain the prediction mechanism of the model. According to SHAP values, the input features are given different weight values to improve the prediction performance.
- Finally, based on the measured spectrum data, the model is evaluated at multiple frequencies. Simulation results demonstrated that the proposed model can obtain preferable interpretability and the accuracy can be further improved as a result of the interpretation.

This paper is organized as follows. In Sec. 2, VMD, LSTM, and SHAP are briefly introduced, and the proposed hybrid interpretable prediction model is constructed. In Sec. 3, the simulation devices and settings are clearly specified, and the evaluation criteria are explained. Section 4 is the simulation test, where we conducted a comparative analysis of the results. Finally, the paper is concluded in Sec. 5.

2. Proposed Model

2.1 Variational Mode Decomposition

The VMD algorithm, proposed by Dragomiretskiy and Zosso [20], provides an adaptive framework for decomposing non-stationary signals into a finite set of IMFs with specific center frequencies. Compared to Empirical Mode Decomposition (EMD), which suffers from mode mixing due to its recursive nature, VMD utilizes an ensemble of adaptive Wiener filters to ensure spectral separation between components. VMD adaptively matches the optimal center frequency and limited bandwidth for each mode, thereby decomposing the original load data into multiple simpler sequences. This process reduces the complexity of prediction and enhances prediction accuracy [11], [12].

Different from the recursive decomposition mode of EMD, the VMD algorithm uses several adaptive Wiener filter banks, the decomposed components can be adaptively segmented. This method effectively overcomes the pattern confusion phenomenon that occurs in the decomposition [18], that is, the same component appears in different frequency bands or different components appear in the same frequency band. The specific implementation steps of VMD are as follows. Firstly, the obtained analytical signal is analyzed after Hilbert transform to calculate the one-sided spectrum. Next, the subsignal is modulated by multiplying the center band to the corresponding baseband. Finally, the demodulated signal gradient norm is calculated to estimate the signal bandwidth of each mode. The constrained variational problem is given by

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

where $u_k = \{u_1, \ldots, u_k\}$ and $w_k = \{\omega_1, \ldots, \omega_k\}$ are the set of each IMF component and the corresponding center frequency, k represents the total number of modes, and $\delta(t)$ represents the Dirac distribution. A Lagrange multiplier operator λ and a quadratic penalty factor α are added for the reconstruction constraints. In the presence of Gaussian noise, α ensures the accuracy of signal reconstruction while maintaining the tightness of the constraints to obtain the optimal solution of the constrained variational problem. Its expression is as

$$\mathcal{L}\left(\{u_k\},\{\omega_k\},\lambda\right) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle. \tag{2}$$

The solution of the minimization problem in (1) is the saddle point of the augmented Lagrange. The Alternate Direction Method of Multipliers (ADMM) is adopted in VMD, and the components and their center frequencies are constantly updated to obtain the optimal solution.

Update process of mode number u_k is as

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum\limits_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}. \tag{3}$$

Update process of center frequency ω_k is as

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left|\hat{u}_k(\omega)\right|^2 \mathrm{d}\omega}{\int_0^\infty \left|\hat{u}_k(\omega)\right|^2 \mathrm{d}\omega}. \tag{4}$$

2.2 Long Short-Term Memory Network

LSTM network is a specialized type of RNNs. It is designed to mitigate the vanishing and exploding gradient issues that occur in traditional RNNs when handling long temporal sequences, and has become one of the most popular RNN architectures [21].

By introducing memory units and gating mechanisms, LSTM can effectively capture long-term dependencies in time series data, which is widely used in natural language processing, time series prediction, speech recognition and other fields. The model structure is shown in Fig. 1.

The core of LSTM is the memory cell, which contains three gating mechanisms.

The forget gate identifies irrelevant information to remove from the memory cell can be expressed as

$$f_t = \sigma(W_\mathrm{f} \cdot [h_{t-1}, x_t] + b_\mathrm{f}). \tag{5}$$

The input gate determines which new information is stored in the memory cell can be expressed as

$$i_t = \sigma(W_\mathrm{i} \cdot [h_{t-1}, x_t] + b_\mathrm{i}), \tag{6}$$

$$\tilde{C}_t = \tanh(W_\mathrm{c} \cdot [h_{t-1}, x_t] + b_\mathrm{c}). \tag{7}$$

The output gate determines what information is to be output from the memory cell can be expressed as

$$o_t = \sigma(W_\mathrm{o} \cdot [h_{t-1}, x_t] + b_\mathrm{o}), \tag{8}$$

$$h_t = o_t * \tanh(C_t). \tag{9}$$

The update of the memory cell state can be expressed as

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{10}$$

where x_t represents the input of the current time step, h_{t-1} represents the hidden state of the previous time step, C_{t-1} represents the memory unit state of the previous time step, W and b represent the learnable weight and bias, σ represents the Sigmoid activation function, and tanh represents the hyperbolic tangent activation function.

2.3 VMD-LSTM Model

Spectral time series has obvious nonlinear and non-stationarity characteristics, and the accuracy of traditional prediction methods is limited. The LSTM neural network demonstrates strong predictive capabilities for low-complexity time series, while the VMD algorithm can effectively address the nonlinearity and non-stationarity of time series data.Therefore, the VMD-LSTM prediction model, which combines the VMD method with the LSTM model, can improve the prediction accuracy. In addition, VMD decomposes the signal into different band of frequency, which is conducive to model interpretation and optimization.

In this work, we propose a hybrid architecture for time series analysis. As shown in Fig. 2, VMD is applied to decompose the spectrum data $X(t)$ into n IMFs and a residual term. IMFS are then fed into parallel LSTM branches for temporal pattern extraction, followed by feature fusion through a fully-connected layer to generate final predictions.

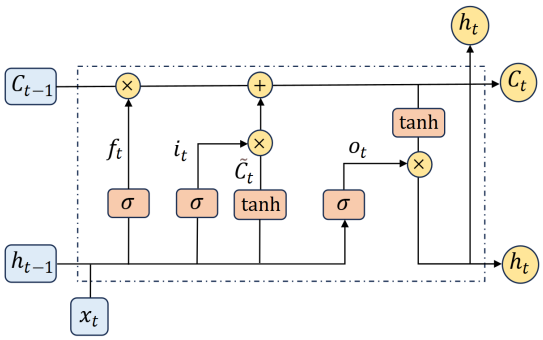


Fig. 1. LSTM structure.

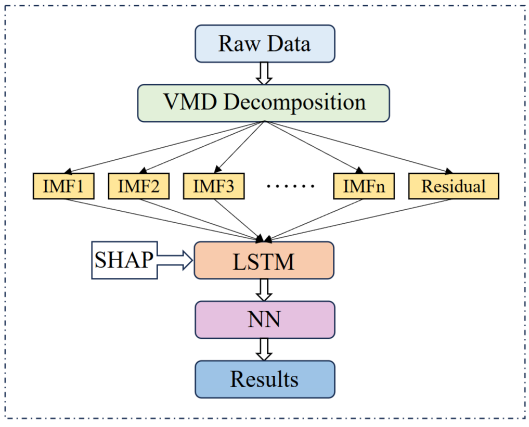


Fig. 2. VMD-LSTM model framework.

2.4 SHapley Additive exPlanations

SHAP is a popular feature attribution mechanism for XAI by computing Shapley values from coalitional game theory to measure the impact of input features on machine learning model [22]. SHAP is based on cooperative game theory, it combines Local Interpretable Model-agnostic Explanations (LIME), Deep Learning Important Features Tracking (DeepLIFT) and other explanation methods from previous researchers. It uses a simplified binary linear function to explain the model [23] as

$$f(x) = g(z) = \phi_0 + \sum_{i=1}^{m} \phi_i z_i \tag{11}$$

where $f(x)$ is the original model, $g(z)$ is the interpretation model, z is the simplified input, $x \rightarrow h_x(z)$is the mapping function to the original method, $\phi_0 = f(h_x(0))$ is the output of the model without simplifying the inputs.

$$\phi_i(f, x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left(f(S \cup \{i\}) - f(S)\right) \tag{12}$$

where, F represents the set of all features, S represents the feature subset, $f(S)$ represents the predicted value of the model using feature subset S, and ϕ_i represents the SHAP value of feature i.

In this paper, the input features are IMFs. By analyzing and calculating the SHAP values of each IMF, we can clarify their impact on the prediction results, thereby achieving interpretability of the model. Compared with previous studies, such as those in [10] and [12], which merely used VMD to improve the model's performance without providing explanations for the model. Our method introduces SHAP to enable the calculation and analysis of the importance of each IMFs. Moreover, we propose an innovative method by incorporating weight coefficients based on the SHAP values during the feature input stage, thereby improving model performance. While existing XAI research predominantly emphasize explanatory analysis, we surpass the explanations by adding a shap-guided weighting scheme to the input sequence, thereby converting theoretical insights into measurable performance improvements.

3. Simulation Settings

The simulation dataset is sourced from the Electrosense open spectrum repository [24], covering the 1300–1764 MHz band with a resolution of 2 MHz. The time-series data consists of 2,661 samples collected at 1-minute intervals, and its time-frequency characteristics are visualized in Fig. 3. The spectrum data was obtained by sensors deployed outdoors in Madrid, Spain, with the frequency set at 1300–1764 MHz, the frequency resolution set at 2000 kHz, and the time interval set at 1 minute. The collected spectrum dataset consists of 2661 time slots and 232 frequency points, representing the signal's signal-to-noise ratio (SNR) in dB. The spectrum data

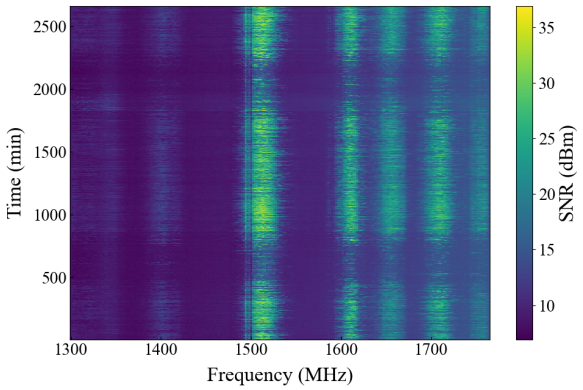


Fig. 3. Time-frequency distribution of spectrum data.

Hyperparameter	Setting
Input sequence length	12
Number of layers	2
First layer units	64
Dropout rate (first layer)	0.2
Second layer units	32
Dropout rate (second layer)	0.1
First dense units	32
Second dense units	1
Optimizer	Adam
Learning rate	0.001
Learning rate schedule	None
Loss function	MSE
Batch size	32
Early stopping rules	Patience = 10
Random seeds	42

Tab. 1. Hyperparameters settings of the model.

is distributed in a band-like pattern with intervals. The SNR in the blue area is relatively low, indicating that it is suitable for communication within this frequency band. The red area has a higher SNR, suggesting that this frequency band is likely to be congested. All simulations were conducted on a workstation equipped with an Intel Core i7-13650HX processor, 16 GB DDR5 RAM, and an NVIDIA GeForce RTX 4060 GPU, utilizing PyCharm 2024.1 with Python 3.10 environment.

RNN [5], LSTM [6], Transformer [7], GRU [8] and WD-LSTM [9] are compared as benchmark models, all adopting a two-layer structure. The neural network architecture is configured with the hyperparameters detailed in Tab.1. Each input sequence comprises 12 historical time steps processed through a two-layer network structure. All layers in the models use the default activation functions.The first layer contains 64 units followed by a dropout layer with a rate of 0.2 to mitigate overfitting. The second layer consists of 32 units with a dropout rate of 0.1. The network further includes two dense layers: the first with 32 units and the final output layer with a single unit using linear activation for regression prediction. The model is optimized using the Adam algorithm with a fixed learning rate of 0.001 and Mean Squared Error (MSE) as the loss function. Training is conducted with a batch size of 32, incorporating early stopping with a patience of 10 epochs. All experiments are conducted

with a fixed random seed of 42 to ensure reproducibility. The VMD-LSTM and WD-LSTM decomposition prediction models differ from other recurrent neural networks primarily in their input data structure, where decomposed components are processed through parallel network branches.

The dataset was partitioned using an 8:2 training-test split. The complete time series data was segmented into sequences of a specified length, which corresponds to the input time steps. The first 80% of the segmented data was used to train the predictive model, ensuring that the model learns from the majority of the dataset. The remaining 20% of the data was reserved for testing the model's performance. Additionally, 20% of the training data was further allocated as a validation set to monitor the model's performance during training and to prevent overfitting.

Figure 4 presents our simulation tests for training epochs to determine the optimal number of training epochs based on the calculation of Weighted Quality Evaluation Index (WQE) [25], which is a novel comprehensive evaluation index to assess the performance of time series forecasting models. These tests were conducted on data sampled at 1500 MHz frequency. As shown in the figure, the model performance gradually improves with the increase in training epochs but eventually stabilizes. It is evident that more training epochs would lead to longer training times. Therefore, considering both model performance and time efficiency, we have chosen to train the model for 70 epochs each time. Additionally, to ensure the stability of the training process and the reliability of the results, each model's training and prediction will be repeated 50 times, with the average criteria metric used as the evaluation result for the model.

Common evaluation metrics include Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), mean absolute error (MAE), and the coefficient of determination (R^2) [25]. RMSE is the square root of the mean of the squared difference between the predicted value and the true value, which is more sensitive to larger errors. MAPE is the average of the relative percentages of the absolute errors between the predicted value and the true value. MAE is used to measure the average degree of difference between the predicted value and the true value. R-squared is a statistical measure used to measure how well the regression model explains the variation in the dependent variable.

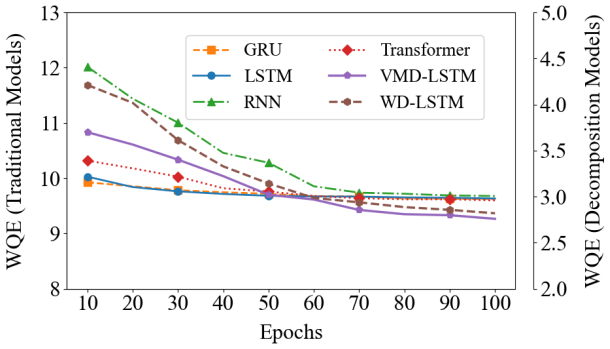


Fig. 4. The evolution of the loss values (WQE) across epochs.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{13}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{14}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{16}$$

where y_i represents the true value, $\hat{y}_i$ represents the predicted value, $\bar{y}$ represents the average of the actual observation values, and n represents the number of samples.

Reference [25] introduced a novel comprehensive evaluation index WQE to assess the performance of global navigation satellite system (GNSS) height time series forecasting models, which can be calculated as

$$\begin{aligned}\mathrm{WQE} = {} & l_1\left(\frac{\mathrm{RMSE}}{X_{\max} - X_{\min}}\right) + l_2\left(\frac{\mathrm{MAE}}{X_{\max} - X_{\min}}\right) \\ & + l_3\mathrm{MAPE} + l_4(1 - R^2)\end{aligned} \tag{17}$$

where $X_{\max}$ and $X_{\min}$ represent the maximum and minimum values of the dataset, respectively. The smaller the WQE, the better the performance. The main advantage of using WQE is that it can perform multi-dimensional performance evaluation, while traditional criteria metrics can only evaluate the prediction performance of the model from a single dimension respectively. By adjusting the value of l , the evaluation of the model can be more biased, accurate and practical, which is helpful for researchers and practical applications to better understand and select suitable prediction models.

The WQE is a comprehensive evaluation metric that combines four different assessment indicators through normalization and weightiing to provide a more holistic evaluation of model performance. The weight parameters l_1 to l_4 are assigned based on the importance of each performance metric in the model evaluation.

The tests shown in Tab.2 are based on the initial settings of $l_1 = 1$, $l_2 = 2$, $l_3 = 0.5$, and $l_4 = 3$, which are computed on the spectrum time series data at a frequency of 1500 MHz. For l_1, while fixing the other three parameters, we calculate the WQE by taking values of 0.5, 1, and 1.5. For l_2, while fixing the other three parameters, we calculate the WQE by taking values of 1, 2, and 3. Similarly, for l_3, we take values of 0.3, 0.5, and 0.8; and for l_4, we take values of 2, 3, and 4. The data in the table show that l_3 has the greatest impact on the results. To emphasize the overall interpretability of the model and the distribution of prediction errors, the highest weight is given to R, followed by the RMSE and MAE. To avoid excessive influence of MAPE on the WQE, it is reasonable to set it to a smaller value.

l_1	WQE	l_2	WQE	l_3	WQE	l_4	WQE
0.5	5.75	1	5.72	0.3	3.92	2	5.56
1.0	5.81	2	5.81	0.5	5.81	3	5.81
1.5	5.88	3	5.91	0.8	8.65	4	6.07

Tab. 2. The impact of different l on WQE.

In order to improve the generalization ability of the model and reduce the distribution range of predicted fluctuations, the parameters are selected as

$$\mathrm{WQE} = \left(\frac{\mathrm{RMSE}}{X_{\max} - X_{\min}}\right) + 2\left(\frac{\mathrm{MAE}}{X_{\max} - X_{\min}}\right) + 0.5\mathrm{MAPE} + 3(1 - R^2). \tag{18}$$

The main evaluation criterion adopted in this study is WQE, supplemented by RMSE and MAPE. In this study, the WQE is dimensionless, the unit of RMSE is dB, which is consistent with the spectral data, and the MAPE is in %.

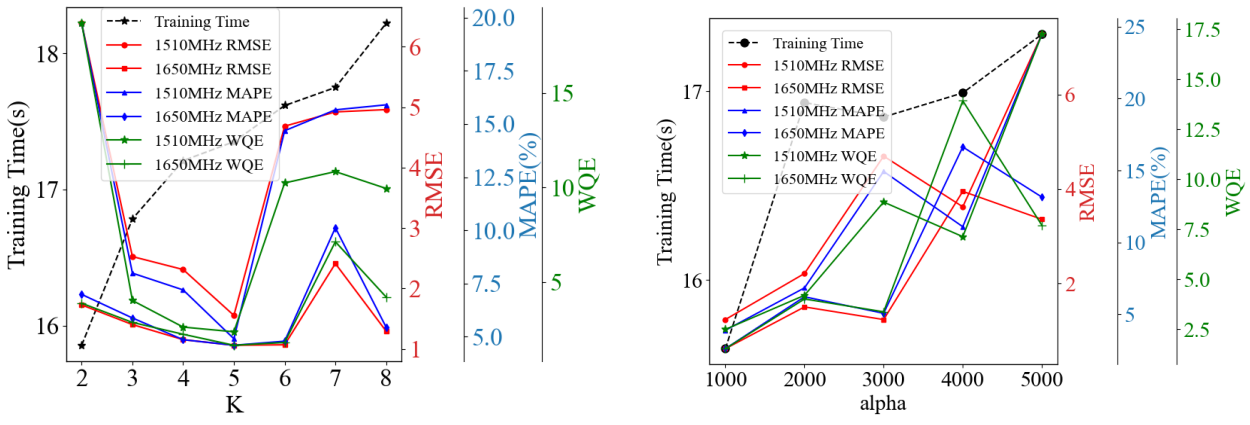
4. Simulation Results and Analysis

4.1 VMD Decomposition Results

VMD involves two critical hyperparameters: the number of IMF (K) and the quadratic penalty factor (α). K is crucial to the decomposition effect of VMD. An inappropriate number of IMFs will lead to the wrong estimation of the center frequency ω_k, and the IMF signal u_k tightly around the center frequency will deviate from the actual signal, resulting in the corresponding mode aliasing phenomenon in the final result. In general, when the value of K is less than the number of resonant frequency bands of the actual input signal, the decomposed IMF will contain multiple resonant frequency bands. When the value of K is greater than the actual number of resonant frequency bands, the actual resonant frequency band will be decomposed into multiple IMF components. α will affect the optimization bandwidth range of IMF in the actual decomposition process, that is, when α is larger, the search range of u_k around the central frequency ω_k is smaller, and the noise introduced is correspondingly reduced. On the contrary, when α is smaller, the search range is larger and more noise is introduced. The simulation examines the impact of these two parameters on the VMD-LSTM model's prediction outcomes, as depicted in Fig. 5. Our parameter selection is based solely on the training data and does not involve the test data.

The test employs data from the 1510 MHz and 1650 MHz frequency points. In the plot, the red line illustrates the trend of RMSE variation with respect to K and α, the blue line depicts the trend of MAPE change, and the black dashed line indicates the relationship between model training time and these two parameters. The results show that when K is set to 5 and α is a relatively small value of 1000, the model exhibits low RMSE and MAPE values, while also requiring a short training time. Thus, taking into account model performance, complexity, and computational cost, K and α are determined to be 5 and 1000, respectively.

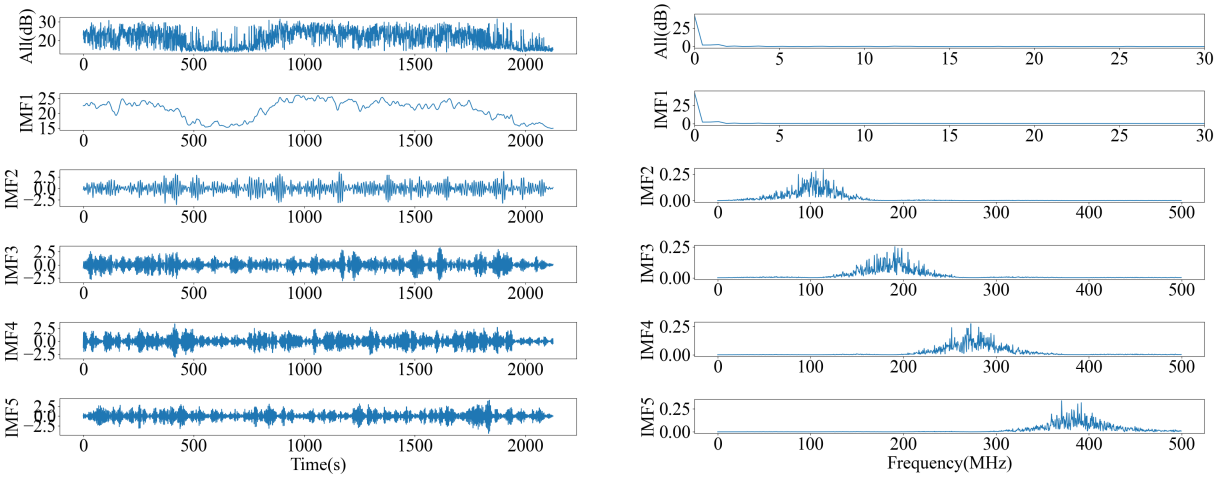
Figure 6(a) shows the VMD decomposition in the time domain, where the sequence is decomposed into five subsequences (IMF1 to IMF5). The vertical axis represents the signal-to-noise ratio (SNR), while the horizontal axis indicates the time points. From top to bottom, the figure displays the original sequence and the five IMFs. Figure 6(b) shows the VMD results in the frequency domain. The subsequences are arranged from low frequency to high frequency, with the horizontal axis representing frequency and the vertical axis representing SNR. As can be seen from the frequency-domain decomposition, the center frequencies of the sub-sequences are separated, which is a significant advantage of VMD over EMD and WD. In Fig. 6(b), IMF1 is a low-frequency sub-sequence that reflects the long-term trend of the spectrum time series, while the other sub-sequences are high-frequency ones, corresponding to rapid and intense data fluctuations. The separation of the center frequency points of each sub-sequence effectively avoids the spectral aliasing issues present in EMD and WD, thus enhancing the model's prediction performance and interpretability.



(a) Number of decomposition modes K

(b) Quadratic penalty factor α

Fig. 5. Effect of hyperparameters of VMD on model performance.



(a) Time domain decomposition

(b) Frequency domain decomposition

Fig. 6. VMD decomposition results.

4.2 VMD-LSTM Model Prediction Results

The VMD-LSTM model was used to predict the spectral data, and 80% of the data was set as the training set. Firstly, the training data was decomposed, and then the results of the decomposition were used to test the VMD decomposition parameters and train the model. Next, the entire sequence was decomposed, and the last 20% of the data in the decomposed sequence was used as the test set. Figure 7 shows the comparison between the prediction results of the model and the real value. It can be seen that the predicted value of the model fits the actual value well.

Tables 3, 4 and 5 present a comparative analysis of the prediction performance between the VMD-LSTM model and the baseline models. In the table, "Freq." refers to "frequency", and "TF" refers to "transformer". It can be seen that, compared to the baseline models, the RMSE and MAPE of the WD-LSTM and VMD-LSTM have significantly decreased, and the prediction performance has been greatly improved. This means that decomposition is beneficial to reduce the sequence complexity and improve the prediction performance. Besides, VMD-LSTM model has better prediction effect than WD-LSTM model.

Table 6 shows the comparison of training time between LSTM model and the two decomposition prediction models. It can be seen that due to the complex structure, the training time of the two decomposition models is longer than that of LSTM model, but the increase of computational complexity and time consumption is not obvious compared with the improvement of model performance. And the training time of the two decomposition prediction models is related to the number of decomposed sub-signals, the more the number of sub-signals, the longer the training time. In short, the VMD decomposition can effectively improve the prediction performance of the model, and the resulting increase in complexity and calculation amount are within the acceptable range.

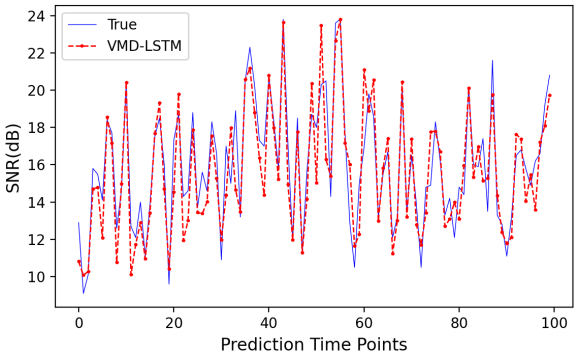


Fig. 7. Comparison of VMD-LSTM prediction results with real values.

Freq. [MHz]	RNN	LSTM	GRU	TF	WD-LSTM	VMD-LSTM
1510	11.64	9.06	9.11	8.91	2.76	2.29
1610	12.51	9.75	9.77	9.75	2.88	2.36
1650	8.66	6.32	6.21	6.17	1.42	1.37
1710	11.35	8.54	8.66	8.46	2.20	3.08
1760	8.32	6.01	6.09	5.87	1.29	1.53

Tab. 3. Performance comparison (WQE).

Freq. [MHz]	RNN	LSTM	GRU	TF	WD-LSTM	VMD-LSTM
1510	5.07	4.18	4.19	4.15	1.72	1.27
1610	4.92	4.02	4.01	4.02	1.70	2.17
1650	2.50	2.06	2.05	2.06	0.84	0.64
1710	4.15	3.39	3.41	3.37	1.35	1.75
1760	1.80	1.44	1.44	1.88	0.84	0.66

Tab. 4. Performance comparison (RMSE).

Freq. [MHz]	RNN	LSTM	GRU	TF	WD-LSTM	VMD-LSTM
1510	16.31	14.84	13.77	13.67	5.45	4.01
1610	17.75	14.98	14.96	15.00	6.06	7.90
1650	10.29	8.67	8.61	8.54	3.41	2.73
1710	14.89	12.55	12.68	12.57	4.88	6.29
1760	8.02	6.51	6.48	7.78	3.41	2.80

Tab. 5. Performance comparison (MAPE%).

Freq. [MHz]	LSTM	WD-LSTM	VMD-LSTM
1510	13.72	14.81	15.74
1610	13.17	14.74	15.16
1650	13.16	14.78	15.15
1710	13.24	14.78	15.57
1760	13.26	14.87	16.97

Tab. 6. Model training time [s].

XAI methods	IMF1	IMF2	IMF3	IMF4	IMF5
SHAP	0.52	0.16	0.04	0.08	0.20
LIME	0.79	0.07	0.03	0.08	0.08

Tab. 7. IMF importance scores derived by XAI methods.

4.3 SHAP Value and Model Improvement

In order to reveal the inherent law of model prediction and improve the model, SHAP is introduced to analyze the VMD-LSTM model. In this study, we selected the KernelExplainer method, which is a sampling and perturbation-based approach that can be applicable to any type of model. To construct the background dataset, we adopted a clustering-based sampling strategy. Specifically, we partitioned the training data into 200 clusters using the k-means algorithm and selected the centroids of these clusters as the background dataset. This strategy effectively maintains the diversity of the data while significantly reducing computational costs, thereby ensuring the efficiency and stability of the SHAP value calculations. When computing the SHAP values, we employed 2048 perturbations, and the random seed was set to 42. The calculation time for the SHAP value is 521.50 s.

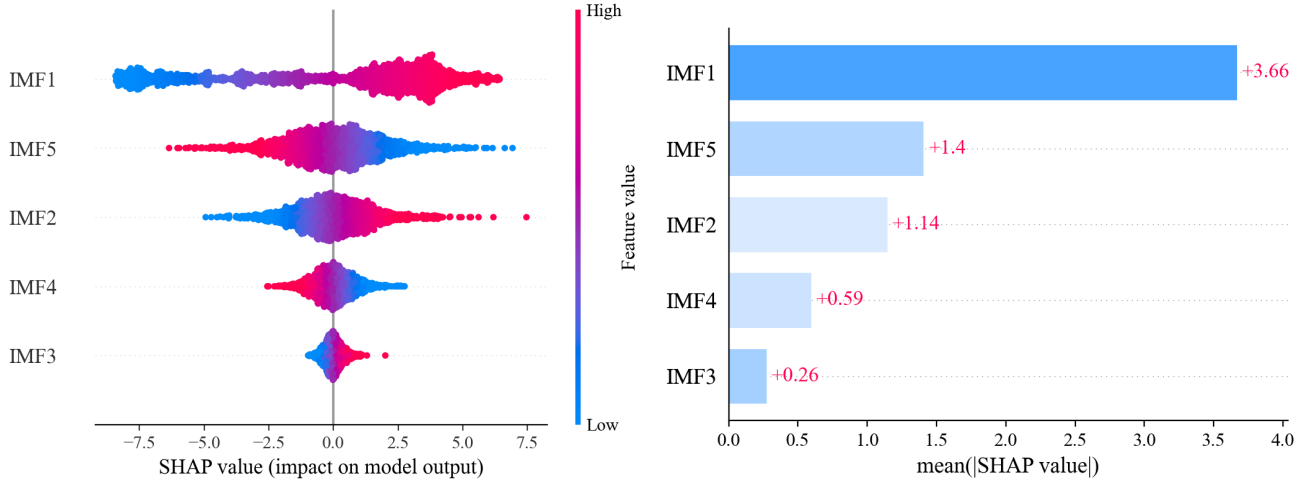
Figure 8(a) shows the distribution of SHAP values for different IMFs, highlighting their varying impacts on each model prediction. Both positive and negative values are present, with larger absolute values indicating greater influence. The SHAP values represent the impact of each feature on the model's output, with IMF1 having the most significant effect. The remaining IMFs, namely IMF5, IMF2, IMF4, and IMF3, show a descending order of importance. Figure 8(b) confirms IMF1's leading role, with the highest average absolute SHAP value, indicating it captures the overall trend of the spectral data, which is crucial for accurate forecasting. Meanwhile, although high-frequency components have a lesser impact on the model's outcomes, they can reveal subtle trends and detailed features in the data and also play a certain role in model prediction, which are associated with the nonlinearity and non-stationarity of the spectral data. The SHAP analysis not only clarifies the model's decision-making process but also enhances its interpretability and performance.

Table 7 compares SHAP with LIME. To facilitate the comparison, the importance scores have been normalized. It can be seen that the LIME [26] method is limited to focusing on the importance of IMF1, which is due to its main analysis of local importance and lack of a global perspective.

After the above calculation and analysis, SHAP value is used as weight to modify the model input as

$$\begin{aligned}\text{Input} = \{&\text{IMF1} \cdot 0.5202 + \text{IMF2} \cdot 0.1614 + \\ &\text{IMF3} \cdot 0.0418 + \text{IMF4} \cdot 0.0778 + \text{IMF5} \cdot 0.1988\}. \end{aligned} \quad (19)$$

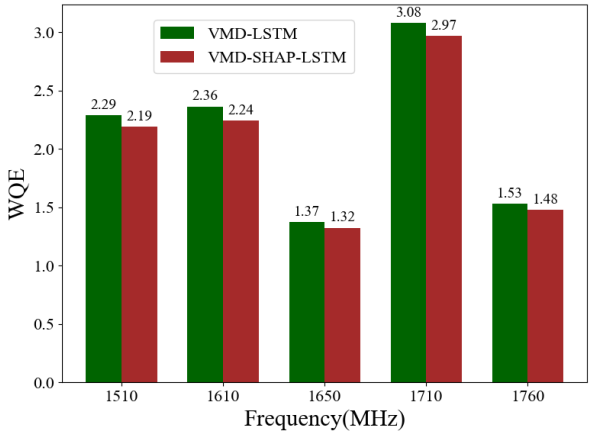
The optimized IMFs data is input into the VMD-LSTM model, and the performance comparison of the optimized model is shown in Fig. 9 and Tabs. 8, 9, 10. In the tables, "Avg." refers to "average". Simulation tests at multiple frequency points show that the method effectively improves the model performance on the basis of explaining the model. The results demonstrate that low-frequency components play a decisive role in spectral time series prediction. While high-frequency components do contribute to the prediction, SHAP analysis enables us to quantify their relative impact through feature weighting, thereby enhancing both the accuracy and reliability of prediction.



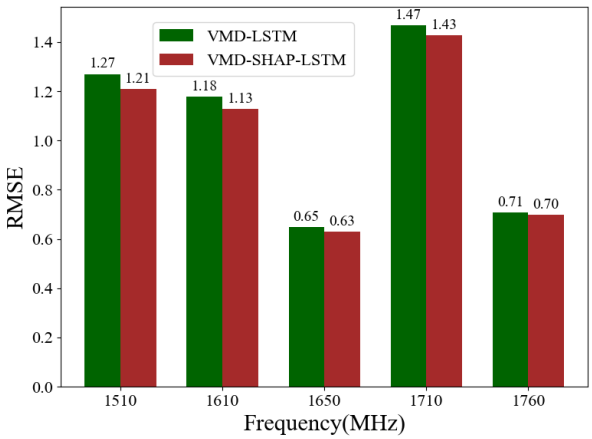
(a) Distribution of SHAP value

(b) Average SHAP absolute value

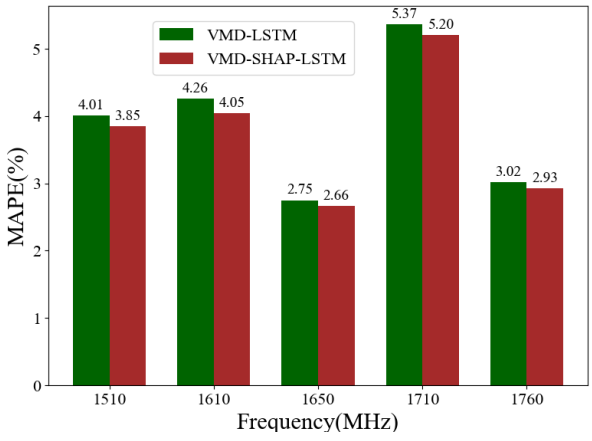
Fig. 8. Results of SHAP value calculation.



(a) Comparison of WQE



(b) Comparison of RMSE



(c) Comparison of MAPE

Fig. 9. Improvement of the Model with SHAP.

Freq. [MHz]	**VMD-LSTM**	**+SHAP**	**WQE%**
1510	2.29	2.19	4.37%
1610	2.36	2.24	5.08%
1650	1.37	1.32	3.65%
1710	3.08	2.97	3.57%
1760	1.53	1.48	3.27%
Avg.	2.13	2.04	3.99%

Tab. 8. Improvement of model with SHAP (WQE).

Freq. [MHz]	**VMD-LSTM**	**+SHAP**	**RMSE%**
1510	1.27	1.21	4.72%
1610	1.18	1.13	4.24%
1650	0.65	0.63	3.08%
1710	1.47	1.43	2.72%
1760	0.71	0.70	1.41%
Avg.	1.06	1.02	3.23%

Tab. 9. Improvement of model with SHAP(RMSE).

Freq. (MHz)	**VMD-LSTM**	**+SHAP**	**MAPE%**
1510	4.01	3.85	3.99%
1610	4.26	4.05	4.93%
1650	2.75	2.66	3.27%
1710	5.37	5.20	3.17%
1760	3.02	2.93	2.98%
Avg.	3.88	3.74	3.67%

Tab. 10. Improvement of model with SHAP (MAPE%).

A detailed statistical comparison between the VMD-LSTM and VMD-LSTM-SHAP models was conducted using Welch's unequal variances t-test [27]. The t-test analysis was performed based on the spectrum data at 1510 MHz frequency. As shown in Tab. 11, the analysis reveals the superior predictive performance of the VMD-LSTM-SHAP model. The WQE, RMSE, and MAPE results of the VMD-LSTM-SHAP model show significant decreases, with t-statistics of 3.517, 3.975, and 3.393, and p-values of 0.00068, 0.00016, and 0.00102, respectively. Figure 10 provides a graphical comparison of the WQE, RMSE, and MAPE between the two models. The VMD-LSTM-SHAP model exhibits lower median errors and tighter error distributions, implying more reliable and accurate forecasting. The 95% confidence intervals for WQE ([1.91%, 6.86%]), RMSE ([2.40%, 7.24%]), and MAPE ([1.67%, 6.39%]) further reinforce the statistical significance of these improvements. Collectively, these findings demonstrate that the VMD-LSTM-SHAP model significantly enhances predictive accuracy compared to the VMD-LSTM model. The standard deviations of the performance metrics for the VMD-LSTM-SHAP model are generally lower than those for the VMD-LSTM model, indicating more stable and consistent performance.

Metric	**WQE**	**RMSE**	**MAPE**
t-statistic	3.517	3.975	3.393
p-value	0.00068	0.00016	0.00102
95% confidence interval	[1.91%, 6.86%]	[2.40%, 7.24%]	[1.67%, 6.39%]
Average improvement	4.37%	4.72%	3.99%
VMD-LSTM std	0.1460	0.0907	0.2204
VMD-LSTM-SHAP std	0.1292	0.0519	0.2095

Tab. 11. Statistical analysis of VMD-LSTM and VMD-SHAP-LSTM.

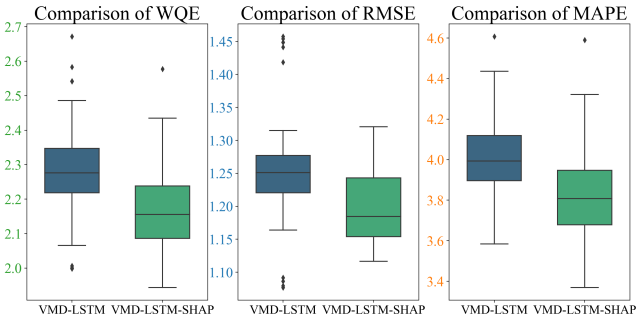


Fig. 10. Box plots of WQE, RMSE and MAPE for VMD-LSTM vs VMD-LSTM-SHAP.

The training time of the improved model matches that of the VMD-LSTM model, but the SHAP method requires additional computation time. In the simulation, calculating the SHAP values for the VMD-LSTM model requires 10 minutes per frequency point. The lengthy computation time of SHAP is primarily due to the multiple iterative perturbations required to accurately estimate Shapley values [28]. To mitigate this, we propose several strategies: Firstly, randomly selecting a subset of the data for SHAP value calculations can achieve preliminary rapid computations at the cost of some precision. Secondly, leveraging parallel processing and hardware acceleration, like using multi-core CPUs or GPUs, can speed up SHAP calculations. These methods help to alleviate the computational burden of SHAP, making it more suitable for real-time or low-latency scenarios.

5. Conclusion

To address the challenges caused by non-stationary and nonlinear spectrum data in existing prediction models, we proposed VMD-LSTM framework, which can improve both prediction accuracy and model interpretability. By using SHAP to quantify the importance of features, our method makes it easier to understand how deep learning models work, which are usually like a black-box. The method improves predictions by calculating SHAP values and subsequently assigning weights to the decomposed signal sequence based on the results. Simulation results demonstrate that the VMD framework significantly enhances model performance, while the SHAP interpretation provides transparent explanations of the decision-making process. Building upon this explanation, we further optimize model performance by incorporating SHAP value-based weighting for inputs. This optimization yields reductions of 4.25% in WQE, 3.77% in RMSE, and 3.84% in MAPE. This method not only improves the prediction accuracy, but also makes the model easier to understand and optimize.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62231012.

References

[1] BECVAR, Z., GESBERT, D., MACH, P., et al. Machine learning-based channel quality prediction in 6G mobile networks. *IEEE Communications Magazine*, 2023, vol. 61, no. 7, p. 106–112. DOI: 10.1109/MCOM.001.2200305

[2] ZHENG, S., CHEN, S., QI, P., et al. Spectrum sensing based on deep learning classification for cognitive radios. *China Communications*, 2020, vol. 17, no. 2, p. 138–148. DOI: 10.23919/jcc.2020.02.012

[3] WANG, L., HU, J., JIANG, D., et al. Deep learning models for spectrum prediction: A review. *IEEE Sensors Journal*, 2024, vol. 24, no. 18, p. 28553–28575. DOI: 10.1109/jsen.2024.3416738

[4] ZUO, P. L., PENG, T., WANG, X., et al. Spectrum prediction for frequency bands with high burstiness: Analysis and method. In *Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. Antwerp (Belgium), 2020, p. 1–7. DOI: 10.1109/VTC2020-Spring48590.2020.9128865

[5] XIA, J., DOU, Z., QI, L., et al. A Hybrid spectrum prediction model based on deep learning. In *Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR)*. Montreal (Canada), 2022, p. 2378–2384. DOI: 10.1109/ICPR56361.2022.9956519

[6] ZUO, P. L., WANG, X., LINGHU, W., et al. Prediction-based spectrum access optimization in cognitive radio networks. In *Proceedings of the 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. Bologna (Italy), 2018, p. 1–7. DOI: 10.1109/PIMRC.2018.8580726

[7] TANG, Y., ZHANG, Y., LI, J. A time series driven model for early sepsis prediction based on transformer module. *BMC Medical Research Methodology*, 2024, vol. 24, p. 1–12. DOI: 10.1186/s12874-023-02138-6

[8] YU, L., GUO, Y., WANG, Q., et al. Spectrum availability prediction for cognitive radio communications: A DCG approach. *IEEE Transactions on Cognitive Communications and Networking*, 2020, vol. 6, no. 2, p. 476–485. DOI: 10.1109/TCCN.2020.2973572

[9] GENG, K., ZHANG, J. Z., YAO, C. H. Spectrum prediction based on wavelet decomposition explainable LSTM. *IEEE Transactions on Consumer Electronics*, 2024, vol. 70, no. 4, p. 7470–7481. DOI: 10.1109/TCE.2024.3433595

[10] ZHAO, L., LI, Z., QU, L., et al. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China. *Ocean Engineering*, 2023, vol. 276, p. 1–18. DOI: 10.1016/j.oceaneng.2023.114136

[11] HAO, Y., LU, J., PENG, G., et al. F10.7 daily forecast using LSTM combined with VMD method. *Space Weather*, 2024, vol. 22, no. 1, p. 1–13. DOI: 10.1029/2023SW003552

[12] LV, L., WU, Z., ZHANG, J., et al. A VMD and LSTM based hybrid model of load forecasting for power grid security. *IEEE Transactions on Industrial Informatics*, 2021, vol. 18, no. 9, p. 6474–6482. DOI: 10.1109/TII.2021.3130237

[13] WU, K., LU, J., LIN, F., et al. A realistic network traffic forecasting method based on VMD and LSTM network. In *IEEE International Symposium on Circuits and Systems (ISCAS)*. Monterey (USA), 2023, p. 1–5. DOI: 10.1109/ISCAS46773.2023.10182143

[14] DING, T., WU, D., SHEN, L., et al. Prediction of significant wave height using a VMD-LSTM-rolling model in the South Sea of China. *Frontiers in Marine Science*, 2024, vol. 11, p. 1–16. DOI: 10.3389/fmars.2024.1382248

[15] RIBEIRO, M. T., SINGH, S., GUESTRIN, C. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco (USA), 2016, p. 1135–1144. DOI: 10.1145/2939672.2939778

[16] ARRIETA, A. B., DIAZ-RODRIGUEZ, N., DEL SER, J., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 2020, vol. 58, p. 82–115. DOI: 10.1016/j.inffus.2019.12.012

[17] JABEUR, S. B., MEFTEH-WALI, S., VIVIANI, J. L. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 2024, vol. 334, no. 1, p. 679–699. DOI: 10.1007/s10479-021-04187-w

[18] WANG, K., ZHANG, L., FU, X. Time series prediction of tunnel boring machine (TBM) performance during excavation using causal explainable artificial intelligence (CX-AI). *Automation in Construction*, 2023, vol. 147, p. 1–18. DOI: 10.1016/j.autcon.2022.104730

[19] MAARIF, M. R., SALEH, A. R., HABIBI, M., et al. Energy usage forecasting model based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI). *Information*, 2023, vol. 14, no. 5, p. 1–18. DOI: 10.3390/info14050265

[20] DRAGOMIRETSKIY, K., ZOSSO, D. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 2013, vol. 62, no. 3, p. 531–544. DOI: 10.1109/TSP.2013.2288675

[21] DIPIETRO, R., HAGER, G. D. Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention*. Zhou, S. K., Rueckert, D., Fichtinger, G. (eds.). Pittsburgh (USA): Academic Press, 2020, p. 503–519. DOI: 10.1016/B978-0-12-816176-0.00026-0

[22] VAN DEN BROECK, G., LYKOV, A., SCHLEICH, M., et al. On the tractability of SHAP explanations. *Journal of Artificial Intelligence Research*, 2022, vol. 74, p. 851–886. DOI: 10.1613/jair.1.13283

[23] ANTWARG, L., MILLER, R. M., SHAPIRA, B., et al. Explaining anomalies detected by autoencoders using Shapley additive explanations. *Expert Systems with Applications*, 2021, vol. 186, p. 1–14. DOI: 10.1016/j.eswa.2021.115736

[24] RAJENDRAN, S., CALVO-PALOMINO, R., FUCHS, M., et al. Electrosense: Open and big spectrum data. *IEEE Communications Magazine*, 2017, vol. 56, no. 1, p. 210–217. DOI: 10.1109/MCOM.2017.1700200

[25] ZHOU, Y., HE, X., MONTILLET, J. P., et al. An improved ICEEMDAN-MPA-GRU model for GNSS height time series prediction with weighted quality evaluation index. *GPS Solutions*, 2025, vol. 29, no. 3, p. 1–19. DOI: 10.1007/s10291-025-01867-z

[26] ROSZYK, N., LEPACZUK, R. The hybrid forecast of S&P 500 volatility ensembled from VIX, GARCH and LSTM models. *arXiv*, 2024, p. 1–38. DOI: 10.48550/arXiv.2407.16780

[27] LU, Z., YUAN, K.-H. Welch's t-Test. In *Encyclopedia of Research Design*. Salkind, N. J. (ed.). Thousand Oaks (USA): Sage Publications, 2010. DOI: 10.13140/RG.2.1.3057.9607

[28] VAN ZYL, C., YE, X., NAIDOO, R. Harnessing eXplainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP. *Applied Energy*, 2024, vol. 353, p. 1–15. DOI: 10.1016/j.apenergy.2023.122079

About the Authors . . .

Wei XU is currently a Graduate Student with the Sixty-Third Research Institute, National University of Defense Technology, Nanjing, China. His main research direction is electromagnetic spectrum management technology.

Jianzhao ZHANG (corresponding author) received the Ph.D. degree in Communication Engineering in 2012. He is currently an Associate Researcher at the Sixty-third Research Institute, National University of Defense Technology, Nanjing, China. His research interests include spectrum environment cognition and smart spectrum management.

Zhe SU received the Ph.D. degree in Communications and Information Systems from the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics in 2023. He is currently a Postdoctoral Researcher in the Sixty-Third Research Institute, National University of Defense Technology, Nanjing, China. His research interests include game theory and communication anti-jamming technology.

Appendix A: Pseudo-Code

Algorithm 1. VMD-LSTM-SHAP framework.

Require: Raw spectrum data X, number of IMFs K, penalty factor α, time steps T

Ensure: Predicted spectrum values $\hat{y}$, SHAP explanations, performance metrics

Phase 1: Data Preprocessing
Load raw spectrum data from Electrosense repository
Normalize data: $X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$
Split data into training set and test set
Phase 2: Variational Mode Decomposition
Initialize VMD parameters: $K = 5$, $\alpha = 1000$, tolerance = $1e^{-7}$
Decompose training data using VMD:
for $i = 1$ to length of training data **do**
    Solve constrained variational problem using ADMM {Eq. (1)}
    Update IMFs {Eq. (3)}
    Update center frequencies {Eq. (4)}
end for
Obtain K IMF components: $IMF_1, IMF_2, \ldots, IMF_K$
Phase 3: VMD-LSTM Model
Construct parallel LSTM branches for IMF processing
Implement feature fusion layer
Train on decomposed IMF components
Phase 4: SHAP Analysis
Initialize SHAP explainer with training data
Calculate SHAP values for each IMF component {Eq. (12)}
Analyze feature importance: $importance_i = \frac{|\phi_i|}{\sum_{j=1}^{K} |\phi_j|}$
Phase 5: Model Improvement with SHAP Weighting
Weight each IMF by its normalized SHAP value: $IMF_i^{\text{weighted}} = IMF_i \times \frac{|\phi_i|}{\max(|\phi|)}$
Retrain LSTM model on weighted IMF data
Phase 6: Prediction and Evaluation
Generate predictions on test set: $\hat{y} = model(IMF_{\text{test}}^{\text{weighted}})$
Calculate performance metrics: WQE {Eq. (18)}, RMSE {Eq. (13)}, MAPE {Eq. (14)}
Output
return Predicted values $\hat{y}$, SHAP values ϕ, performance metrics