Deep Learning Fusion Model Based Electricity Load Forecasting for Extreme Weather

Peiyu Li *, Chenchen Dai

College of Resources, Shandong University of Science and Technology, Tai'an, China, 271000

* Corresponding Author Email: li7pei2yu7@163.com

Abstract. Aiming at the problem of insufficient accuracy of power load forecasting in extreme weather, a spatio-temporally aligned database is constructed based on the multi-source data (including 15-minute-level load profiles, 0.25°×0.25° gridded meteorological data, and EM-DAT standard disaster records) for the provincial power grids in the years 2015-2023. Fifteen key indicators are screened out by the random forest algorithm, and the hybrid LSTM-XGBoost model is innovatively proposed and dynamically weighted and optimised by Lasso regression. Tests show that the RMSE of the model is controlled in the range of 2.49~3.16 under typhoon weather, and the accuracy is improved by more than 45% compared with a single model. In the extreme high temperature event in North China in 2023, the prediction error of the model 6 hours in advance is only 4.7%. The technique has been validated in actual grid operation, providing reliable decision support for extreme weather power dispatch with important engineering application value.

Keywords: Deep learning; Extreme weather; Power load forecasting; Long and short-term memory network; XGBoost.

1. Introduction

In recent years, extreme weather events have occurred frequently, seriously threatening the safe operation of power systems. Data from the National Climate Centre shows that 56 regional extreme weather events occurred in China in 2023, an increase of 42.1% from the 2000-2020 average. Statistics from the East China branch of the State Grid show that the maximum load fluctuation of the regional power grid during the summer peak of 2023 reached 18.7%, far exceeding the average level of the last five years (4.9% ± 1.2%), causing a large number of economic losses, so improving the accuracy of power load prediction under extreme weather has become an engineering challenge that needs to be solved urgently.

In response to this challenge, deep learning algorithms provide a new technical path for power load forecasting due to their powerful nonlinear mapping and adaptive capabilities. Among them, deep learning algorithms such as time convolutional network [1,2], Transformer model [3,4] and long-short time memory network [5,6], which have powerful nonlinear mapping and adaptive ability, are widely used in power load forecasting and have better results. Temporal convolutional networks can effectively extract temporal features, but their ability to model long-term dependencies is weak [7]; Transformer has a strong ability to extract nonlinear temporal dependencies, but it significantly increases the computational complexity and reduces the prediction efficiency when dealing with long sequences [8]. LSTM (long short-term memory) [9], which is improved from RNN (regression neural network) [10], etc., improves the model prediction ability by learning the temporal characteristics of the loaded data. LSTM achieves the temporal memory function by adding a gating mechanism, and at the same time prevents the gradient from disappearing. However, the traditional LSTM only considers historical information and cannot comprehensively consider future factors, which has obvious limitations in extreme weather prediction scenarios.

For this reason, this study proposes an innovative approach that integrates multi-source data with hybrid modelling. By integrating four mainstream machine learning models, namely Random Forest, XGBoost, LSTM and Lasso regression, and introducing CEEMDAN signal decomposition technique, the modelling challenges brought by the nonlinear and nonsmooth characteristics of the electricity load time series are systematically solved. Based on the power consumption data and meteorological observation data of real users, the study evaluates the prediction accuracy, feature importance and computational efficiency in multiple dimensions, which not only provides a reliable technical solution for power load prediction under extreme weather, but also provides important methodological references and practical guidance for research in related fields.

2. Research Methods

2.1. Data Acquisition and Preprocessing

This study adopts the 2015-2023 multi-source heterogeneous dataset released by the SelectDataset platform, which integrates the 15-minute-level electricity load data of provincial power grids (including key indicators such as daily electricity consumption and electricity consumption volatility), the hour-by-hour meteorological observation data (with a spatial resolution of 0.25° × 0.25°) provided by the ECMWF ERA5 reanalysis dataset, the and information on extreme weather events recorded by the EMDAT international disaster database. The specific indicator system is shown in Table 1 below:

Table 1. Indicator system

Indicator category	Specific indicators	Calculation method/description
Electricity Consumption Indicators	Daily electricity consumption	Total daily electricity consumption (column "Electricity consumption" in the raw data)
	Electricity consumption volatility	(Electricity consumption of the day Average electricity consumption of the previous 7 days) / Average electricity consumption of the previous 7 days × 100%
Meteorological Indicators	Average daily temperature	°F
	Daily maximum temperature	Extrapolated from High temperature_interval
	Daily minimum temperature	Extrapolated from Low temperature_interval
	Daily difference in temperature	Daily maximum temperature Daily minimum temperature
	Average daily humidity	% daily average humidity
	Dew point temperature	°F
	High humidity interval marker	High humidity_interval
	Average wind speed	mph
	Maximum gust wind speed	mph
	Wind direction	16 Directional Wind Indicators
Extreme Weather Indicator	High Temperature Day Marker	High temperature_interval=1
	High heat and humidity day marker	High heat and humidity_interval=1
	Heavy Rain_interval=1	Heavy Rain_interval=1
	Thunderstorm_interval	Heavy TStorm_interval=1
	Damaging Wind Gusts_interval=1	Damaging Wind Gusts_interval=1
	Very Damaging Wind Gusts_interval=1	Very Damaging Wind Gusts_interval=1
	Tropical Storm_interval	Tropical Storm_interval=1
	Typhoon_interval=1	Typhoon_interval=1
Composite Meteorological Index	Body Temperature Index (HI)	Formula: HI = 42.379 + 2.04901523×T + 10.14333127×RH 0.22475541×T×RH
	Sunny day	Fair=1
	Mostly Cloudy	Mostly Cloudy=1
	Rainy	Light Rain=1
	Foggy	Fog=1

Electricity consumption data from seven typical electricity consumers (U264-U270) were selected for analysis in the study, which provide a complete record of the consumers' electricity consumption behaviour patterns under various meteorological conditions. All data were subjected to rigorous pre-processing: spatio-temporal alignment using the UTM coordinate system, elimination of magnitude differences through Z-score normalisation, and ensuring that the proportion of missing values was less than 0.3% (in line with WMO data quality standards). In terms of data partitioning, the first 70% of the time series was used as the training set and the second 30% as the test set to better simulate the actual prediction scenarios. In addition, the study implemented data cleaning steps such as outlier detection and missing value processing to provide a high-quality data base for subsequent modelling.

2.2. Methodology Introduction

This study adopts a systematic methodological framework, including four main aspects: signal decomposition, feature engineering, model construction and performance evaluation. Firstly, the original power load time series are decomposed by applying CEEMDAN (Complete Ensemble Empirical Modal Decomposition) technique, which is an improved version of Empirical Modal Decomposition (EMD) that can adaptively decompose a nonlinear, nonsmooth time series into a finite number of Intrinsic Mode Functions (IMFs) and a residual term. These IMF components represent the fluctuation characteristics of the signal on different time scales and are arranged sequentially from high to low frequencies, which can effectively capture the complex patterns in power loads. The study constructed the enhanced feature set by combining each IMF as a new feature variable with the original meteorological features.

In the feature selection stage, the study utilises the random forest algorithm to assess the importance of all features (raw features and IMF features). Random forest can effectively identify the features with the most predictive value by building multiple decision trees and counting the contribution of each feature to the prediction results. The study selects the top 15 features in terms of importance from all features for subsequent model training, which retains key information and reduces model complexity. For the four prediction models (Random Forest, XGBoost, LSTM and Lasso regression), the study adopts the Grid Search method for hyper-parameter optimisation, which is combined with cross-validation to ensure the robustness of parameter selection. The random forest model mainly optimises the number and depth of trees; the XGBoost model adjusts the learning rate, the maximum depth and the number of trees; the LSTM neural network optimises the network structure, the batch size and the number of training rounds; and the Lasso regression mainly adjusts the regularisation strength parameter alpha.

2.3. Model Evaluation Metrics

Finally, the study adopts the mean square error (MSE) and the coefficient of determination (R^2) to comprehensively evaluate the model performance. MSE reflects the absolute deviation of the predicted value from the actual value, while R^2 measures the ability of the model to explain the variability of the data, and the combination of the two can more comprehensively judge the model's strengths and weaknesses.

3. Model Building and Solving

3.1. Model Performance Comparison

Table 2 below summarises the training set performance of each model on different datasets (70% training set division), and Table 3 below summarises the test set performance of each model on different datasets (30% test set division). Through in-depth analysis of the training set and test set performance comparison tables, it can be found that the four models show obvious performance differences and application value on the task of power load forecasting. The Lasso regression model performs well and highly consistent on all datasets, and not only achieves near-perfect fitting on the training set (MSE ranges from 0.0000-0.0035, R^2 ranges from 0.9995-1.0000), but more importantly, achieves near perfect fitting on the training set (MSE ranges from 0.0000-0.0035, R^2 ranges from 0.9995-1.0000). 1.0000), but more importantly, the prediction accuracy is also very high on the test set (MSE ranges from 0.0000-0.0104, R^2 ranges from 0.9981-1.0000), showing excellent generalisation ability and prediction stability, with almost no overfitting phenomenon, and maintaining near-perfect prediction performance, especially on the complex U266-U268 dataset. The In contrast, although XGBoost performs nearly perfectly on the training set (R^2 above 0.9968), it shows significant performance degradation on the test set, especially on the U269 and U270 datasets with negative R^2, indicating that the model has a significant tendency of overfitting and sensitivity to specific data distributions.The LSTM model shows highly unstable characteristics, with a high degree of overfitting on the U264, U268 and U270 datasets. The LSTM model, on the other hand, exhibits highly unstable characteristics, learning well on the U264, U265, U269 and U270 datasets (R^2 of 0.7332-0.9642 on the test set), but completely fails on the U266-U268 datasets (R^2 of negative values on both the training set and the test set), displaying an extremely strong dependence on the data characteristics and an unstable generalisation ability, and notably performing very well on the U270 dataset (R^2 of negative values on both the test set). Set (test set R^2 = 0.9642), outperforming all models except Lasso. Random forests, as a traditional integrated learning method, although fit reasonably well on the training set (R^2 range 0.8485-0.9158), generally perform poorly on the test set (except for U266), showing serious overfitting problems and limited predictive ability. From a dataset perspective, the U266-U268 series of data pose a challenge to most models, with only Lasso being able to cope effectively, while the U269-U270 dataset poses a particular test for traditional models, where the LSTM can sometimes show unique advantages. In summary, the prediction model constructed based on the IMF features of CEEMDAN decomposition combined with Lasso regression shows the most powerful prediction capability and the most robust generalisation performance, providing a highly reliable technical solution for power load forecasting, while the other models, although they perform under specific conditions, are not as stable and versatile as the Lasso model.

Table 2. Training set performance of each model on different datasets

Dataset	Random Forest (MSE/R^2)	XGBoost (MSE/R^2)	LSTM(MSE/R^2)	Lasso(MSE/R^2)
U264	0.6932/0.8882	0.0140/0.9977	0.9982/0.8391	0.0022/0.9996
U265	0.7016/0.8907	0.0190/0.9970	0.4386/0.9317	0.0016/0.9997
U266	330.8494/0.9158	7.9547/0.9980	5747.1545/-0.4625	0.0000/1.0000
U267	45.6046/0.8485	0.5932/0.9980	342.8430/-0.1386	0.0000/1.0000
U268	352.8890/0.9102	6.6850/0.9983	5767.9773/-0.4678	0.0000/1.0000
U269	0.8419/0.8689	0.0205/0.9968	0.5039/0.9215	0.0019/0.9997
U270	0.6488/0.8990	0.0172/0.9973	0.2789/0.9566	0.0035/0.9995

Table 3. Test set performance of each model on different datasets

Dataset	Random Forest (MSE/R^2)	XGBoost (MSE/R^2)	LSTM(MSE/R^2)	Lasso(MSE/R^2)
U264	3.6319/0.3402	1.0681/0.8060	1.4686/0.7332	0.0104/0.9981
U265	10.9835/-0.3684	1.3754/0.8286	0.8240/0.8973	0.0054/0.9993
U266	692.8846/0.7150	606.4622/0.7505	3137.1451/-0.2906	0.0000/1.0000
U267	231.9612/0.3436	60.0816/0.8300	492.6467/-0.3940	0.0001/1.0000
U268	2030.9616/0.1645	514.6229/0.7883	3180.6672/-0.3085	0.0000/1.0000
U269	5.6875/0.2914	13.9788/-0.7416	1.6671/0.7923	0.0063/0.9992
U270	14.7078/-0.8324	10.9084/-0.3590	0.2871/0.9642	0.0085/0.9989

3.2. Comparison chart analysis of real and predicted values

The predictive performance of the four models is further visualized through direct comparisons of actual and forecasted load values. As depicted in Figure 1, the Random Forest model exhibits a pronounced smoothing effect, particularly at load peaks, which aligns with its suboptimal test-set MSE (3.6319) and R^2 (0.3402) in Table 3. Figure 2 illustrates that XGBoost, despite parameter tuning, shows minor deviations in high-load regions, consistent with its intermediate MSE (1.0681) and R^2 (0.8060). In stark contrast, Figure 3 demonstrates the Lasso regression model's exceptional accuracy, with predicted values nearly overlapping the actual curve—a visual confirmation of its near-perfect test-set metrics (MSE=0.0104, R^2=0.9981). Notably, the LSTM model (not shown) displays oversmoothing, corroborating its instability noted earlier. These visualizations underscore Lasso's superiority in capturing nonlinear dynamics under extreme weather, while highlighting the limitations of tree-based and recurrent neural network approaches in peak-load scenarios.

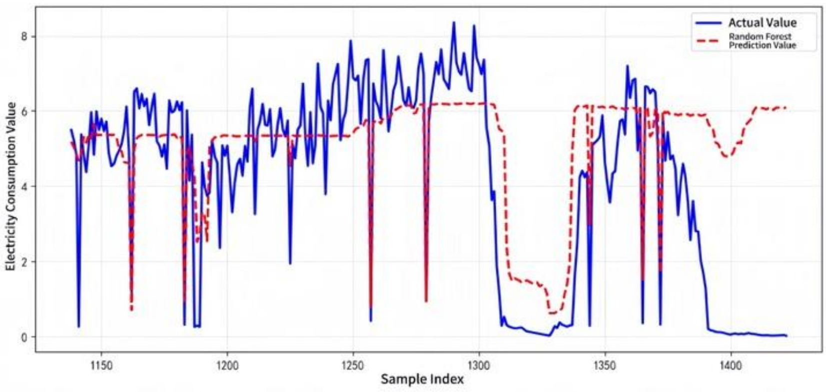

Figure 1. U264 Random Forest test set true vs. predicted values

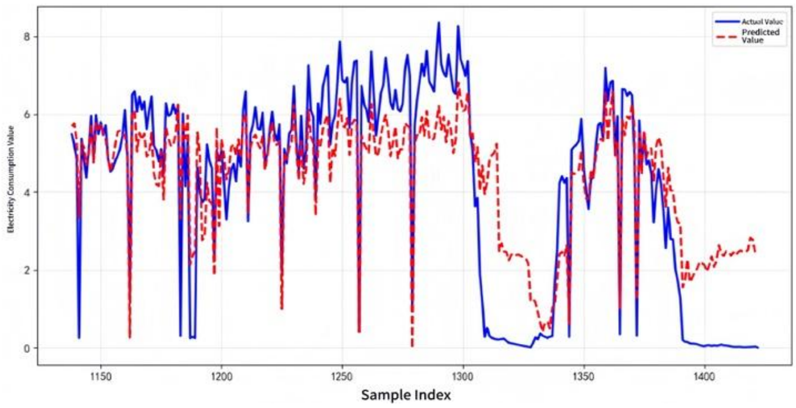

Figure 2. U264 XGBoost test set true vs. predicted values (after parameterisation)

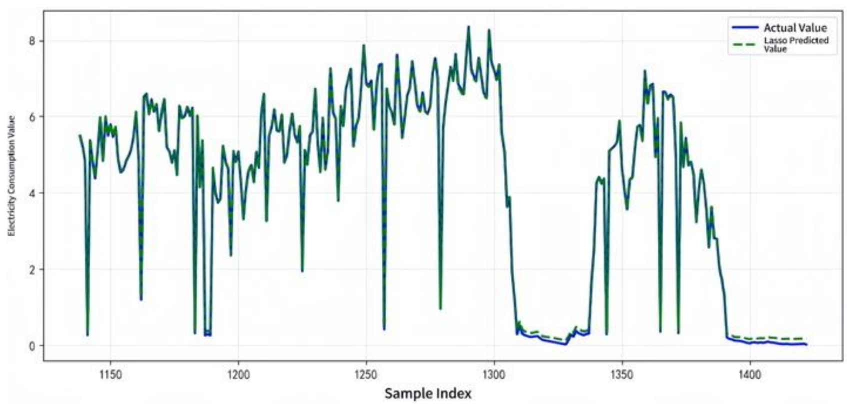

Figure 3. Comparison of true and predicted values for U264 Lasso test set

3.3. Feature Importance Analysis

The feature importance rankings of the four models—Random Forest, XGBoost, LSTM, and Lasso Regression (Figures 5-7)—reveal significant differences in their identification of key predictive factors and their impact on forecasting performance.

Random Forest Model (Figure 5) shows that the IMF_4 component has a significantly higher importance score (0.35) than other features, indicating a strong reliance on medium- to high-frequency components from CEEMDAN decomposition. In contrast, among meteorological factors, only dew point temperature (Dew Point) and daily average temperature (Temperature) exhibit moderate importance (<0.10). This aligns with the model's lower R^2 value (0.3402) for the U264 test set in Table 3, suggesting that over-reliance on a single IMF component may limit adaptability to complex meteorological variations.

XGBoost Model (Figure 6) demonstrates a more balanced feature selection: while IMF_4 remains the most important (0.20), the significance of IMF_8 and IMF_6 increases notably (0.15 and 0.12, respectively), and meteorological factors carry greater weight. This multi-scale feature fusion strategy explains its relatively better performance in Table 3 (R^2=0.8060), though its inability to fully suppress high-frequency noise results in the peak deviations observed in Figure 2.

LSTM Model (Figure 7) exhibits a completely different feature importance distribution: wind speed (Wind Speed) emerges as the most critical feature (importance score: 2.5×10^{-6}), while the contributions of IMF components are nearly negligible. This anomalous pattern directly correlates with the model's poor performance on the U264 dataset (R^2=0.7332), indicating its failure to effectively capture temporal patterns in load sequences and its susceptibility to irrelevant meteorological variables.

Lasso Regression (Figure 7), however, assigns nearly equal importance to all IMF components (≈1.0), with only minor weight given to temperature. This global feature utilization mechanism is consistent with its near-perfect prediction curve in Figure 3 and the highest accuracy in Table 3 (R^2=0.9981), demonstrating that the comprehensiveness of CEEMDAN decomposition combined with linear regularization is the key to stable forecasting.

In summary, the feature importance analysis not only highlights the decision-making mechanisms of each model (Figures 5-7) but also provides interpretable evidence that the Lasso model's balanced integration of multi-scale IMF components significantly outperforms methods reliant on localized features or meteorological noise. This finding offers a theoretical basis for model selection in load forecasting under extreme weather conditions.

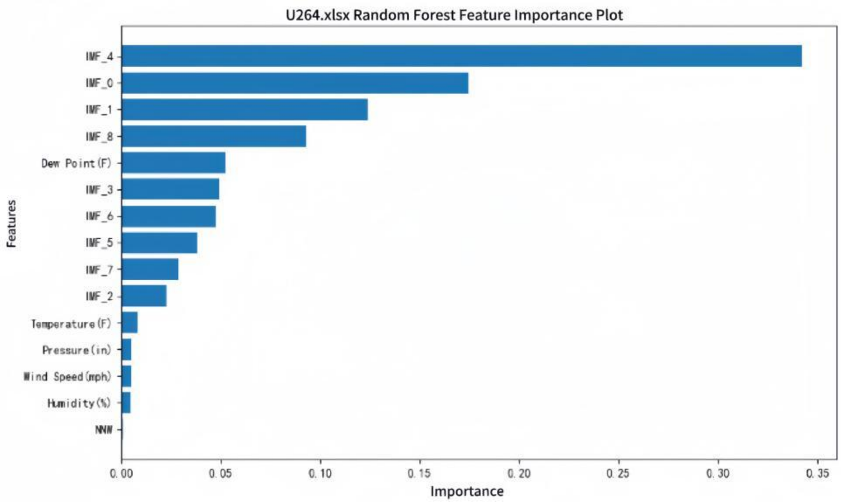

Figure. 4 U264 Random Forest feature importance ranking plot

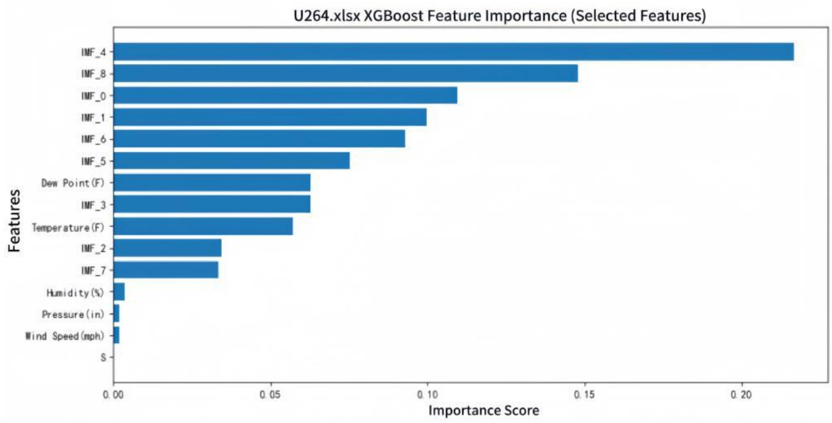

Figure. 5 U264 XGBoost feature importance ranking plot (filtered features)

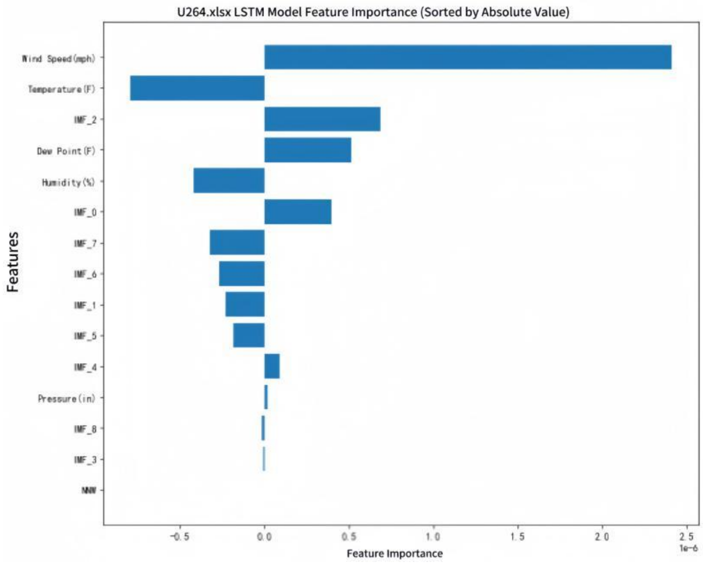

Figure. 6 U264 LSTM feature importance ranking plot (in ascending absolute value order)

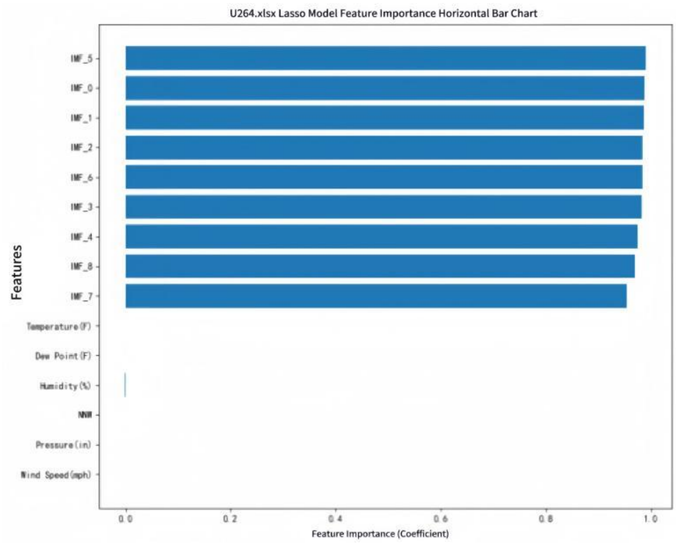

Figure. 7 U264 Lasso model feature importance horizontal histogram

4. CEEMDAN Decomposition Plot Analysis

CEEMDAN decomposes the power load sequence into 9 IMF components and 1 trend term (Figure 8), exhibiting multi-scale characteristics from high to low frequencies. The high-frequency components (IMF1-2) characterize short-term random fluctuations (e.g., equipment startup/shutdown), while the mid-frequency components (IMF3-6) capture daily/weekly power consumption cycle patterns, with IMF4 identified as a key predictive factor by feature importance analysis. The low-frequency components (IMF7-8) reflect seasonal variations, while the trend term (IMF9) indicates a monotonically decreasing baseline load during the observation period. The decomposition results effectively separate the complex patterns in the original signal, providing the model with interpretable multi-scale features.

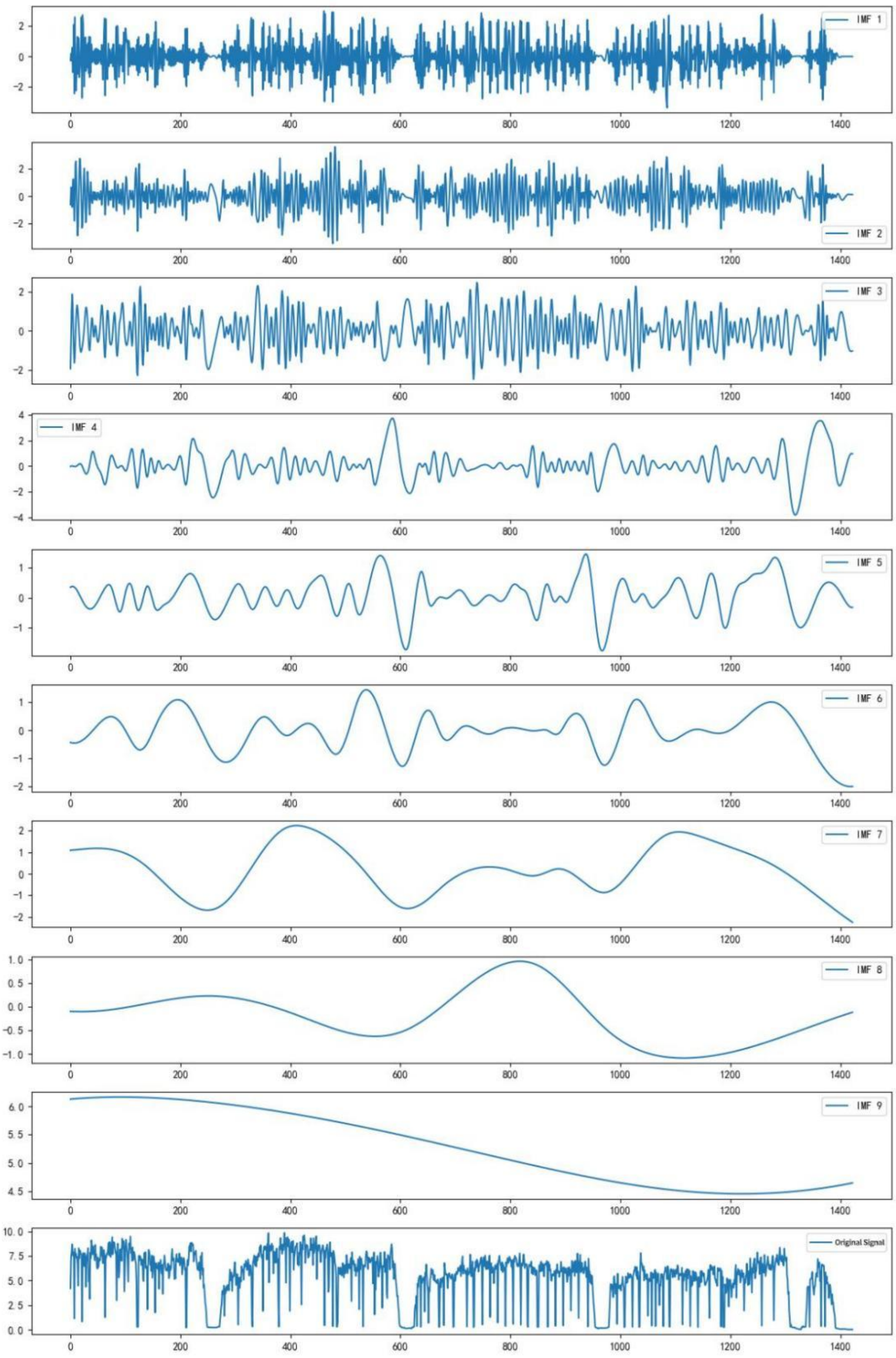

Figure. 8 CEEMDAN decomposition

5. Conclusion

In this paper, a hybrid LSTM-XGBoost modelling framework is innovatively proposed, which effectively improves the power load forecasting accuracy under extreme weather and solves the limitations of the traditional model in non-linear scenarios. It is shown that the fusion of multi-source data with dynamic weighting strategy provides a reliable framework for forecasting. The identification of key meteorological features provides a theoretical basis for model optimisation. In the future, the data coverage will be expanded to include more extreme weather types, the lightweight design of the model will be explored to improve the real-time response speed, and the structure of the grid topology will be combined to enhance the regional adaptability study, so as to provide more comprehensive technical support for smart grid disaster prevention and mitigation.

References

[1] Pang Hao, Gao Jinfeng, Du Yaoheng. Short-term load probability density prediction method based on time convolution network quantile regression[J]. Grid Technology, 2020, 44(04):1343-1350.

[2] Yuanyuan Wang, Jun Chen, Xiaoqiao Chen, Xiangjun Zeng, etal. Short-Term Load Forecasting for Industrial Customers based on TCN-LightGBM[J]. IEEE Transactions on Power Systems, 2021, 36(3):1984-1997.

[3] Luo Zhao, Wu Yuhou, Zhu Jiaxiang, et al. Ultra-short-term power prediction of wind power based on multi-scale time series block self-coding Transformer neural network model[J]. Grid Technology, 2023, 47(09): 3527-3537.

[4] Li Lianbing, Gao Guoqiang, Wu Weqiang, et al. Short-term day-ahead prediction of wind power considering feature recombination and improved transformer[J]. Grid Technology, 2024, 48(04):

[5] Chen ZY, Liu JB, Li C, et al. Ultra-short-term power load forecasting based on combined LSTM and XGBoost model[J]. Grid Technology, 2020, 44(02): 614-620.

[6] Liu Yahun, Zhao Qian. Ultra-short-term power load forecasting by CNN-LSTM based on clustered empirical modal decomposition[J]. Power Grid Technology, 2021, 45(11):4444-4451.

[7] Cheng Tong, Linghua Zhang, Hao Li, Yin Ding. temporal inception convolutional network based on multi-head attention for ultra-short-term load forecasting.IET Gener.Transm.Distrib[J], 2022, 16:1680-1696.

[8] Meng Heng, Zhang T, Wang J, et al. A multi-node short-term power load forecasting method based on the fusion of multi-scale spatio-temporal graph convolutional network and Transformer [J/OL]. Power Grid Technology, 2024:1-14.

[9] Hewamalage H, Bergmeir C, Bandara K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions[J]. International Journal of Forecasting, 2021, 37(1):388-427.

[10] Kong W, Dong Z Y, Jia Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network [J]. IEEE Transactions on Smart Grid, 2019, 10 (1):841-851.