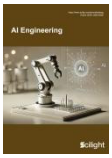
AI Engineering
https://www.sciltp.com/journals/aieng



Article

On Wide-Band Oscillation Localization in Power Transmission Grids: Explainability and Improvement

Yuyou Li, Jie Gu *, Jiaqing Wu, Zhijian Jin and Honglin Wen

Key Laboratory of Control of Power Transmission and Conversion, Department of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Minhang District, Shanghai 200240, China
* Correspondence: gujie@sjtu.edu.cn

How To Cite: Li, Y.; Gu, J.; Wu, J.; et al. On Wide-Band Oscillation Localization in Power Transmission Grids: Explainability and Improvement. *AI Engineering* **2025**, *1*(1), 2.

Received: 14 Apr 2025
Revised: 23 May 2025
Accepted: 30 May 2025
Published: 13 June 2025

Abstract: The broadband oscillation caused by the large-scale integration of new energy generation units into the power grid poses hidden dangers to the stable operation of the power grid. Fast and accurate positioning of the oscillation source is the basis for cutting off the oscillation source. In order to improve the interpretability and accuracy of the broadband oscillation positioning model, this paper proposes an interpretability framework for the transmission network broadband oscillation positioning model, mainly including the improved broadband oscillation model and its interpretation framework. This model integrates graph convolutional neural network and long short-term memory network, takes transmission network measurement sampling data as input, and establishes a broadband oscillation localization model in a data-driven manner; An explanatory framework was constructed for the proposed wideband oscillation localization model, which combines global and local interpretations based on the additive interpretation of Shapley values to improve the interpretability of the wideband oscillation localization model. Based on the explanatory results, an attention feature mechanism is introduced into the localization model to enhance the wideband oscillation localization model. This article uses MATLAB/Simulink (version 2024b) to build a power grid model, produces a sample dataset, and verifies the feasibility and effectiveness of the proposed explanatory framework through numerical simulation.

Keywords: broadband oscillation; explainable framework; SHapley Additive exPlanations; global interpretation; attention feature

1. Introduction

In recent years, with the large-scale integration of new energy power generation units represented by wind power and photovoltaic power into the grid, the penetration level of power electronic devices in power systems has been continuously increasing. Although this has improved power supply efficiency and promoted the clean energy transition, the wide-frequency oscillations caused by the interaction between power electronic devices and the grid pose significant risks to the safe and stable operation of the power system [1]. To suppress wide-frequency oscillations, it is necessary to quickly determine the location of the oscillation source [2] and remove it to achieve the goal of oscillation suppression.

Currently, wide-frequency oscillation localization can be accomplished through two primary methods: mechanism-based models and data-driven models. Mechanism-based models require consideration of dynamic modeling of electronic devices, coupling and transfer characteristics between devices, and source-end aggregation effects, making them complex and generally requiring assumptions and simplifications of the system. On the other hand, data-driven wide-frequency oscillation models have the advantage of strong learning capabilities for

nonlinear complex relationships among massive data and rapid adaptability to random time-varying environments. They can extract effective feature information from measured data to achieve nonlinear fitting between inputs and outputs, thereby completing oscillation source localization. For example, literature [3] uses ensemble learning methods for oscillation source localization, taking into account measurement data errors and changes in power system operating conditions. Literature [4] estimates confidence levels for individual buses and selects the D-S evidence theory as the oscillation source localization algorithm, which is also applicable in multi-oscillation source scenarios. However, the opacity of such models has attracted widespread attention, and the lack of interpretability of black-box models affects the persuasiveness of analysis results and greatly hinders the practical application of these methods in engineering [5]. At the same time, when information is missing or there is a large amount of false information, the effectiveness of machine learning models significantly declines, making it impossible to provide favorable reference results and even leading to serious consequences [6].

Interpretability refers to that the decision-making process and operational logic within an artificial intelligence model align with the current level of human knowledge, and the operational rules of the model itself can be understood and flexibly applied by humans [7]. The stronger the interpretability of a model, the more persuasive it is and the easier it is for people to accept [8]. Machine learning models are complex, with many layers making input-to-output information transfer slow and cumbersome. Detailed exploration is labor-intensive and often impractical. Instead, the focus should be on ensuring users trust the decision-making process and results. Interpretability can be categorized as pre-modeling and post-modeling [9].

Classic pre-modeling interpretability algorithms include decision trees, logistic regression, and naive Bayes models. The common feature of these algorithms is that the model structure itself is simple and easy to understand, meaning the model itself has a certain level of interpretability. However, since the structure of these models is relatively simple, their fitting ability is also limited, making it difficult to handle complex tasks.

Post-modeling interpretability primarily targets inherently non-interpretable models, using additional techniques to make them more acceptable to users [10]. Common post-modeling methods include sensitivity analysis [11], surrogate models [12], and SHapley Additive exPlanations (SHAP) [13]. Sensitivity analysis, a local interpretability method, examines how model outputs change with fine-tuned inputs to quantify feature influence. Literature [14,15] proposed improved sensitivity analysis methods and verified them with examples in the field of image recognition. Literature [16] have explored model output changes by removing feature subsets.

The essence of surrogate models is to select an interpretable model to approximate the original model for the purpose of analogy-based explanation. Literature [10] trained a new interpretable model based on the target model and completed the approximation and explanation of the target model. Literature [8] proposed a local interpretability method by establishing a linear surrogate model to explain artificial neural network models.

The idea of the SHAP value method originates from game theory. In machine learning, SHAP values can characterize the contribution proportion of each feature to the model's output results, thereby explaining the decision-making process of machine learning models. Literature [17] used SHAP values to interpret extreme gradient boosting models, analyzing the degree of influence of features on the results. Literature [18] interpreted Catboost models based on SHAP values and improved model parameters based on the interpretation results.

This paper proposes an interpretability framework for wide-frequency oscillation localization models in transmission networks using SHAP values. The framework calculates SHAP values for each feature to explain the model's decision-making process globally and locally. It then ranks features by importance, removes low-contribution features to simplify data and reduce complexity. A feature attention mechanism is also introduced to enhance model accuracy and efficiency. The framework's effectiveness is validated using data from a four-machine two-area simulation model, showing its practical value for wide-frequency oscillation localization.

This paper constructs a wide-frequency oscillation localization model for transmission grids. Among them, Graph Convolutional Neural Network (GCN) is utilized for capturing spatial dependencies between nodes in the graph, which is crucial for understanding the relationships among different entities in our dataset. In detail, GCN generates global oscillation information using graph convolution and deconvolution layers, based on power grid topology and partial node measurements. It addresses missing data issues through compressive sensing and noise filtering, completes node data, eliminates measurement gaps, adapts to grid topologies, and captures spatial propagation patterns of oscillation energy.

Long Short-Term Memory (LSTM) network is incorporated to model temporal dependencies, allowing the model to capture changes over time. In detail, LSTM which takes the oscillation data of each node as input, captures the time-varying characteristics of broadband oscillations through its recurrent structure. LSTM can effectively handle the spatio-temporal propagation characteristics of broadband oscillations and improve the ability to identify complex oscillation patterns.

The SHAP technique enhances model interpretability by quantifying feature contributions—globally identifying key factors in localization and locally revealing the model's reasoning (e.g., highlighting nodes near oscillation sources in Figures 5 and 6). Feature optimization is achieved by sorting SHAP values to prune low-contribution inputs, reducing data by ~20% and improving efficiency. Complementing this, the feature attention mechanism dynamically adjusts input weights to prioritize critical features, transforming raw vectors into weighted representations that amplify key oscillation signals. This dual approach yields a 1.7% accuracy gain (97.3% → 99%), suppresses noise from secondary features, and strengthens robustness.

Finally, the attention mechanism is introduced to focus on the most relevant information at each time step, enhancing the model's ability to make accurate predictions.

The main contributions of this paper are as follows:

(1) Proposing an interpretability framework based on SHAP values: Provides global and local explanations for wide-frequency oscillation localization models, enhancing the transparency and credibility of the model's decision-making process.
(2) Analyzing feature importance and simplification: Ranks features by importance using SHAP values, eliminates features with low contributions, and reduces data dimensionality and computational complexity.
(3) Introducing a feature attention mechanism: Improves the wide-frequency oscillation localization model by focusing more on key features during training, significantly enhancing the model's localization accuracy.

2. Transmission Network Oscillation Localization Model via Graph Convolution and LSTM

Based on the oscillation signals collected from the transmission network, this paper establishes a wide-frequency oscillation localization model that considers the topological structure of the transmission network. A global oscillation information generation model is constructed using GCN to obtain oscillation data for unknown nodes based on limited measurement data. Then, the measurement data from all nodes in the network are integrated to form an oscillation feature matrix. Considering the time-series characteristics of measurement data during wide-frequency oscillations in the power grid, a LSTM artificial neural network is selected to build the wide-frequency oscillation localization model, achieving oscillation source localization across different frequency bands.

2.1. Global Oscillation Information Generation Model Based on GCN

2.1.1. Graph Signal Model of the Transmission Network

As is known from graph theory, the graph network of the transmission network can be represented as follows:

$$G = (V, E) \tag{1}$$

Among them, $V = (v_1, v_2, \cdots v_N)$ represents the set of nodes, E represents the set of connecting lines, and N represents the number of nodes. Assuming that only M nodes' data are known in the entire network, the voltage and current magnitudes, phase angles, active power, and reactive power of the M nodes can be obtained respectively. That is, the information matrix X formed by the characteristics of each node in the transmission network is:

$$X = \left\{ X_{ua}, X_{up}, X_{va}, X_{vp}, X_{ap}, X_{rp}, \right\} \tag{2}$$

Among them, X_{ua}, X_{up}, X_{va}, X_{vp}, X_{ap} and X_{rp} denote the graph signals for voltage magnitude, voltage phase angle, current magnitude, current phase angle, active power, and reactive power, respectively.

For simplicity, this paper numbers the nodes sequentially based on observability. Nodes 1 to M have observable information, while nodes M + 1 to N have unknown information (i.e., null graph signals). Node observability is affected by incomplete measurement device coverage and asynchronous measurement data.

2.1.2. Global Oscillation Information Generation Model

The global oscillation information generation model takes two inputs: the node data information matrix X and the topology graph adjacency matrix A, where X contains oscillation data for each node. To enhance the efficiency of the wide-frequency oscillation localization model, a compressed sensing algorithm is used. This algorithm retains the essential oscillation characteristics of the original data while suppressing noise. Consequently, the compressed oscillation data is used to construct the node information matrix X, with unobservable nodes' data set to null values.

The node information matrix X is fed into the graph convolution layer for convolutional encoding to extract features from the limited oscillation data, with results stored in an intermediate hidden layer matrix. Subsequently, multiple layers of graph deconvolution restore these aggregated features back to the initial feature space. To mitigate noise amplification during deconvolution, a graph signal noise filtering layer is added, reducing interference in the global oscillation signal. This process ultimately achieves global observability by generating comprehensive global oscillation information.

The global oscillation information generation model leverages the topological connections between nodes and the partial oscillation data collected from observable nodes to supplement missing data due to communication delays. The extracted features from this process serve as the input for the wide-frequency oscillation localization model. The detailed framework is illustrated in Figure 1.

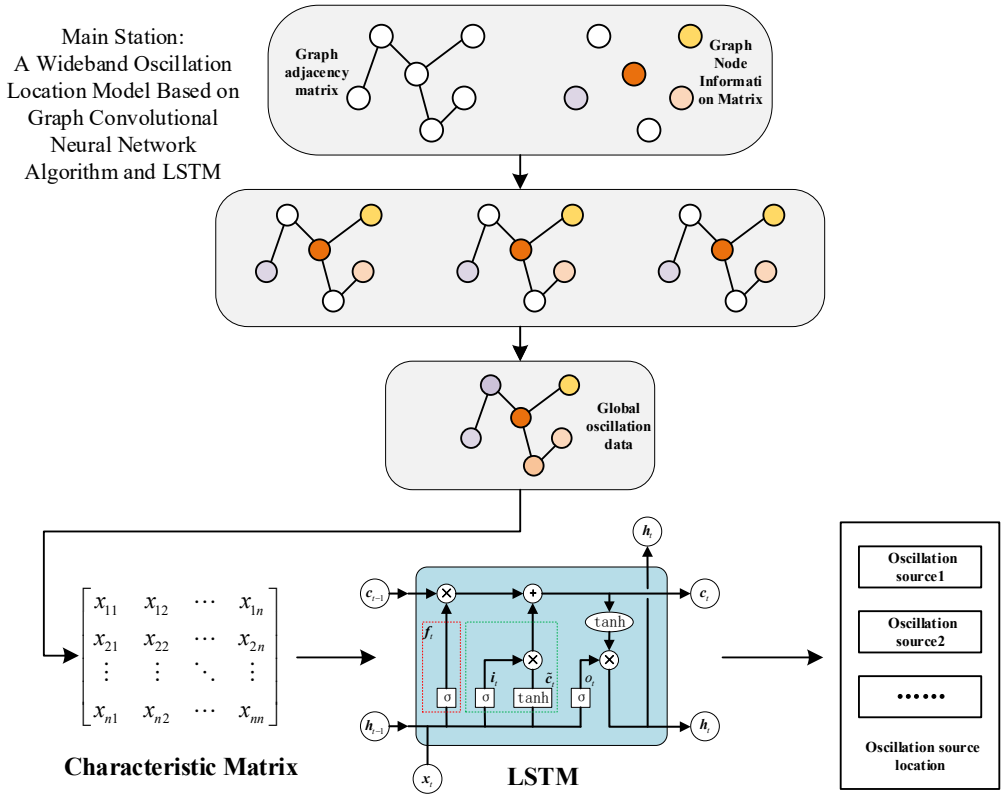


Figure 1. Framework for Broadband Oscillation Localization Using Graph Convolutional LSTM Network.

2.2. Wide-Frequency Oscillation Localization Model Based on LSTM Network

Assume that the number of units in a transmission network is P, and each unit is represented by S_1, S_2 … S_P. The relationship between the location of the oscillating unit and the system measurements can be expressed as:

$$S = F(R) = F(O(Y_{in})) \tag{3}$$

Here, S denotes the oscillating unit, R represents the wide-frequency oscillation characteristics, and Y_{in} is the global measurement input to the model. $F(g)$ indicates the relationship between wide-frequency oscillation characteristics and the oscillation source, while $O(g)$ shows the relationship between input measurements and wide-frequency oscillation characteristics. Meanwhile, the global measurements can be expressed as:

$$Y_{in} = f(Z_{part}) \tag{4}$$

Here, Z_{part} represents the observable system measurement data, and $f(g)$ denotes the functional relationship fitted by the global oscillation information generation model based on GCN. In large-scale power grids, wide-frequency oscillations often exhibit spatiotemporal distribution characteristics, with the propagation path influenced by the oscillation source location and correlations between nodes along the path. As a powerful tool for time series processing, the LSTM network can effectively classify sequences of node measurement information for oscillation source localization.

In this paper, the feature matrix composed of the compressed measurement data of each node is used as the input matrix for the LSTM network, and its expression is:

$$X_{in} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1b} \\ x_{21} & x_{22} & \cdots & x_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ x_{a1} & x_{a2} & \cdots & x_{ab} \end{bmatrix} \tag{5}$$

Here, a is the dimension of the compressed measurement data for a single electrical quantity, which is also the time step input to the LSTM model. b is the total number of electrical quantities in the network, M is the total number of nodes, and d is the number of electrical quantities collected at each node.

3. Interpretability Framework

The wide-frequency oscillation localization model can effectively identify oscillation sources in power systems. However, due to the 'black box' nature of deep learning models, they lack transparency and interpretability, hindering their adoption in practical applications. To address this, this paper introduces SHAP values and attention mechanisms to enhance the model's interpretability and improve its generalization capabilities.

3.1. SHAP Values

The essence of SHAP values is to calculate the marginal contribution of features to the model's output, refining the model's output into a linear function of binary variables. Its expression is as follows:

$$g(x') = \phi_0 + \sum_{i=1}^{Q} \phi_i x_i' \tag{6}$$

Here, ϕ_0 denotes the baseline value of the output result, Q represents the number of features in the localization model, ϕ_i represents the SHAP value of the i^{th} feature, and $x' \in \{0,1\}^Q$ represents the Boolean variable of the i^{th} input feature, indicating whether the i^{th} feature exists in the data sample. The total number of features M defines the dimensionality of the input data, which is processed by the LSTM network at each time step. Together, these elements provide a comprehensive framework for understanding the model's decision-making process in oscillation source localization.

From Equation (6), it can be seen that for a certain oscillation source, whenever the model adds a new feature input, it will affect the final output probability value of the model. Here, the probability value refers to the final probability that the unit is the oscillation source, and the change in the probability value ΔP is the SHAP value of that feature. Therefore, by calculating the SHAP values of all features, one can intuitively perceive the degree of influence of the model's input features on the output value.

3.2. SHAP-Based Interpretability of the Oscillation Localization Model

For the wide-frequency oscillation localization model, assume that for the k^{th} unit, the model's output value under the i^{th} sample is $y_{i,k}$, the feature sequence of the i^{th} sample is $x_{i,k}$, and the j^{th} feature is represented by $x_{ij,k}$. The baseline output value of the localization model is $y_{base,k}$, so the SHAP value satisfies the following expression:

$$y_{i,k} = y_{base,k} + f(x_{i1,k}) + f(x_{i2,k}) + \cdots + f(x_{in,k}) \tag{7}$$

Here, $f(x_{ij,k})$ represents the SHAP value of the feature $x_{ij,k}$, n denotes the number of features, $f(x_{i1,k})$ represents the contribution proportion of the 1st feature to the output probability value of the localization model under the i^{th} sample. The SHAP value quantifies the degree of influence of adding this feature on the change in the output value. If $f(x_{ij,k}) > 0$, it indicates that adding this feature increases the probability value of the unit being the oscillation source, producing a positive effect; conversely, it indicates that adding this feature reduces the probability value.

Calculating SHAP values for different features can enhance the interpretability of the wide-frequency oscillation localization model. This is done from two perspectives:

(1) Local Interpretation: Compute SHAP values for each feature in a single sample to understand its specific impact on the model output. This detailed analysis forms the basis for decision-making.
(2) Global Interpretation: Aggregate SHAP values across multiple or all samples to assess feature importance at a data-driven, variable-level.

3.3. Improving the Wide-Frequency Oscillation Localization Model via SHAP Values and Attention Mechanism

SHAP values quantify the impact of input features on the model's output probability. By ranking features based on their SHAP values, the model can be improved through feature selection: retaining significant features and discarding less important ones, thereby reducing model complexity and computational costs.

Additionally, since all features have equal weight coefficients in the model, it hampers the model's ability to quickly identify core oscillation features. To address this, an attention mechanism is introduced to dynamically adjust feature weights, enhancing the model's focus on important features.

To optimize the wide-frequency oscillation localization model, comprehensive SHAP values for different features are calculated for each oscillation source. This involves computing the weighted average of SHAP values across various oscillation units. Interpretation framework of broadband oscillation positioning model based on SHAP values is shown in Figure 2.

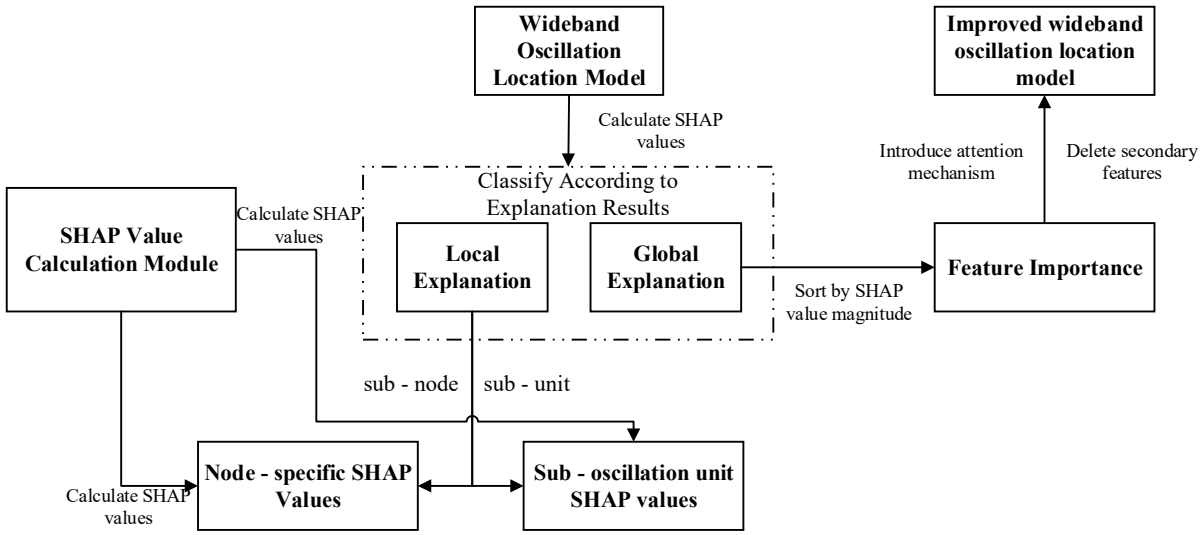


Figure 2. Interpretation framework of broadband oscillation positioning model based on SHAP values.

3.3.1. Feature Attention Mechanism

The core idea of the feature attention mechanism is to continuously adjust the attention weights of input features during the model training process, exploring the contribution proportion of input features to the target features. It directs more attention to key features while diminishing focus on secondary features. Taking the LSTM network model in this paper as an example, when inputting to the t^{th} time step, the input feature vector x_t at this moment can be represented as $x_t = [x_{1,t}, x_{2,t}, \cdots, x_{Q,t}]$, where Q represents the number of features. The attention weight vector x_t can be obtained through a single-layer neural network $e_t = [e_{1,t}, e_{2,t}, \cdots, e_{Q,t}]$, and its calculation formula is as follows:

$$e_t = \sigma(W_e x_t + b_e) \tag{8}$$

Here, σ represents the activation function, e_t denotes the weight vector, $a_t = [a_{1,t}, a_{2,t}, \cdots, a_{m,t}, \cdots, a_{Q,t}]$ represents the feature attention weights. By passing the weight vector e_t through the Softmax layer, the feature attention weights $a_t = [a_{1,t}, a_{2,t}, \cdots, a_{m,t}, \cdots, a_{Q,t}]$ can be obtained. Combining the attention weight vector with the input feature vector results in a weighted vector $x_t^{'}$ adjusted by the weight coefficients. This weighted vector can serve as the new input feature for the LSTM network model, and its calculation expression is as follows:

$$x_t^{'} = a_t * x_t = [a_{1,t}x_{1,t} \quad a_{2,t}x_{2,t} \quad \cdots \quad a_{Q,t}x_{Q,t}] \tag{9}$$

3.3.2. Improvement of the LSTM Network Model

By incorporating the feature attention mechanism discussed earlier into the wide-frequency oscillation localization model, an enhanced LSTM network—termed Feature Attention LSTM (FALSTM)—is developed. This improved model introduces a feature attention layer to the original LSTM architecture. The specific implementation steps are as follows:

(1) Feature Attention Layer: Apply the feature attention mechanism to the input feature matrix Y_{in} to dynamically allocate attention weights to each feature based on their correlations. Compute the actual input feature matrix $Y_{in}^{'}$ by combining these weights with Y_{in}.

(2) LSTM Network Layer: Construct the LSTM layer and obtain the hidden states of each memory unit after $Y_{in}^{'}$ processing the input through the LSTM network.

(3) Fully Connected and Softmax Layers: Feed the hidden states, which contain oscillation information, into a fully connected layer. Normalize the outputs through a Softmax layer to obtain the probability values for each unit being the oscillation source.

(4) Output Layer: Compare the probability values of each unit and identify the unit with the highest probability as the oscillation source, outputting its label.

4. Case Study

4.1. Case Data Overview

To verify the effectiveness of the wide-frequency oscillation localization model proposed in this paper, a four-machine two-area simulation model with a direct-drive wind turbine is built on the MATLAB/Simulink platform based on literature [19]. Node 5 in the model is the equivalent replacement of a 300 MW direct-drive wind farm, and the schematic diagram of the model is shown in Figure 3.

By adjusting the control parameters of the wind turbine in the model, the system can be induced to experience sub-synchronous oscillations, super-synchronous oscillations, and medium-high frequency oscillations. At the same time, periodic disturbance signals with different amplitudes and frequencies are injected at the prime mover ends of the four synchronous generators or the grid-side controllers of the direct-drive wind turbines. The amplitude ranges from 0.1 to 1.6 V with a step size of 0.1 V, and the frequency ranges are 10–40 Hz, 60–90 Hz, and 130–160 Hz with a step size of 0.1 Hz. Additionally, the load levels at L_1 and L_2 are changed (90%–110%, with a step size of 10%) to ensure that the simulated operating conditions in this paper cover a wider sample space of wide-frequency oscillations.

The voltage of each node, the current of each line, active power, and reactive power are sampled at a sampling frequency of 4.8 kHz, and random noise is added to form the original wide-frequency oscillation dataset. Each data sample is correspondingly labeled with the oscillation source location, constructing the data sample library for this case study. A Four Machine Two Area Simulation Model with Direct Drive Fans is shown in Figure 3.

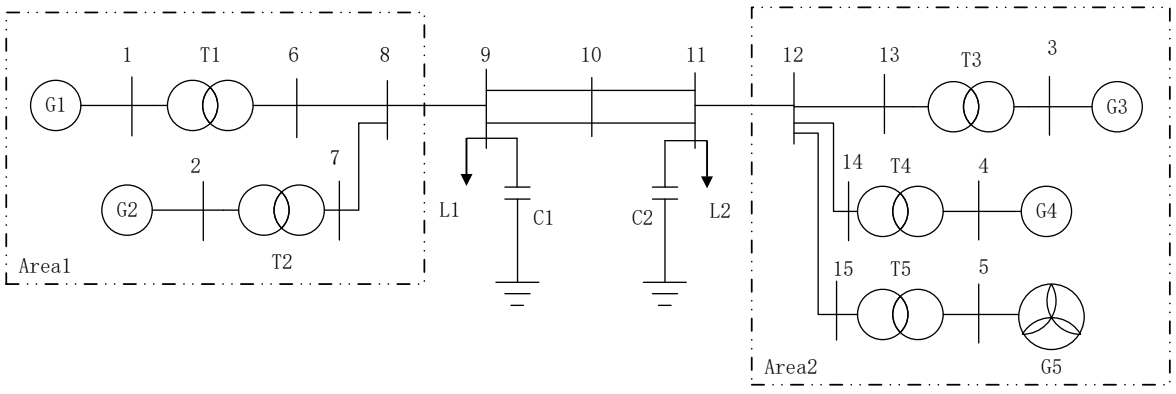


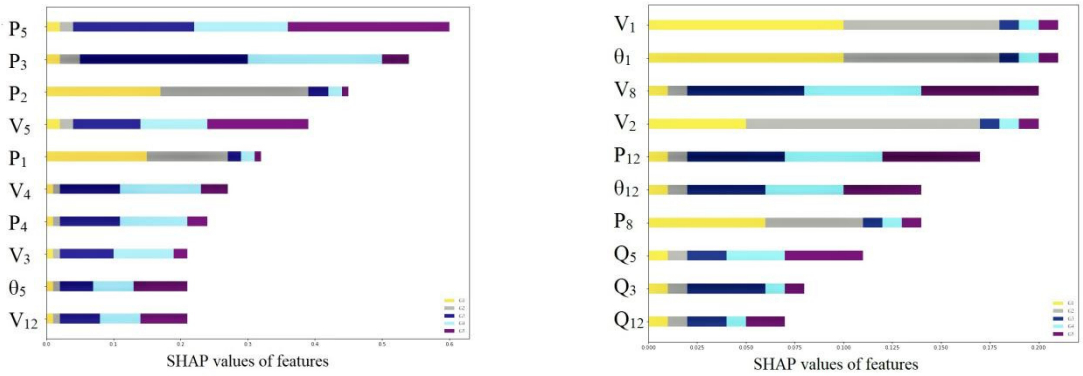
Figure 3. A Four Machine Two Area Simulation Model with Direct Drive Fans.

4.2. Global Interpretation of the Wide-Frequency Oscillation Localization Model

The input feature matrix of the wide-frequency oscillation localization model proposed in this paper has a dimension of 120 × 90, where 120 denotes the number of time steps, and 90 is the dimension of the input feature vector at each step. This characterizes the total number of features for 15 nodes, with each node having 6 feature

quantities (voltage magnitude, voltage phase angle, current magnitude, current phase angle, active power, and reactive power).

The SHAP value method calculates the SHAP values for each feature across all oscillation source units, quantifying their contributions to the output. The weighted average SHAP values are ranked in descending order. Due to space constraints, only the top 20 features with the highest SHAP values are summarized and shown in Figure 4a,b.



(**a**) (**b**)

Figure 4. (**a**) SHAP values of the top 10 quantities. (**b**) SHAP values for the 11th to 20th feature quantities.

P_5 represents the active power of Bus 5, V_5 represents the voltage magnitude of Bus 5, and θ_5 represents the voltage phase angle of Bus 5, and others are expressed similarly. The horizontal axis indicates SHAP values, and five colors represent the labels of the five oscillation sources, highlighting each feature's varying contributions to different oscillation units.

From Figure 4, the following conclusions can be drawn:

(1) Global SHAP Analysis: The active power at Bus 5 has the highest contribution, indicating that the active power measurement at the wind turbine outlet significantly impacts the accuracy of wide-frequency oscillation localization. When the oscillation source is the wind turbine unit, the localization model's accuracy reaches 99%.

(2) Top 10 SHAP Values: The SHAP values of active power from Bus 1 to 5 are among the top 10, highlighting that active power at generator unit outlets is crucial for wide-frequency oscillation localization accuracy.

(3) Oscillation Source Contribution: SHAP values at the oscillation source outlet are higher than those at other nodes. For example, when the oscillation source is G5, active power, voltage magnitude, and voltage phase angle at Bus 5 have the highest contributions. This suggests that the model's decision-making for identifying the oscillation source primarily relies on feature quantities at the oscillation unit outlet.

By analyzing Figure 4a,b, the following conclusions are drawn:

(1) Feature Contribution Ranking: The 20th highest SHAP value is the reactive power at Bus 12, with a total SHAP value of 0.07 across the 5 oscillation sources, averaging 0.014 per source. This indicates that only a few features significantly impact the wide-frequency oscillation localization model, while many others have minimal influence on localization accuracy.

(2) Key Node Analysis: Besides Bus 1–5, Bus 8 and Bus 12 also have relatively high SHAP values. These nodes are key connection points for G1–G2 and G3–G5 in the transmission network and are core nodes of the four-machine, two-area system. Extracting features from these nodes helps accurately identify the oscillation source region, aiding final localization decisions.

(3) Current Measurements: The top 20 SHAP values exclude current measurements, suggesting that current signals contribute less to oscillation localization. While current signals still contain oscillation features, power and voltage signals are easier for deep learning models to extract. Thus, current measurements can be selectively included in the input features based on need.

4.3. Node-Type Interpretation of the Wide-Frequency Oscillation Localization Model

Based on the global interpretation of the wide-frequency oscillation localization model, this section provides a node-type interpretation of the localization model, aiming to study and analyze the varying degrees of influence of measurement data from different nodes on the final output value of the model. From the simulation case

diagram, the node types in the four-machine two-area system can be roughly divided into: generator outlets, generator connections to the transmission network, and the transmission network side, as detailed in Table 1.

Table 1. Example model node type.

Node Type	Generator Outlet	Generator Connection to Transmission Network	Transmission Network Side
Number of Nodes	1, 2, 3, 4, 5	8, 12	9, 10, 11

Since the four synchronous generators in the model share similar internal structures and operating principles, this section compares the model analysis results when the oscillation sources are synchronous generators (G1) and wind turbines (G5). The SHAP values of corresponding features are calculated, and the SHAP values of the six features at each node are summed. The wide-frequency oscillation localization model is then interpreted from the perspective of node types, with results shown in Figures 5 and 6.

In the figures, "base value" represents the baseline value, which can be obtained by averaging all samples, and '*f(x)*' represents the average probability value finally obtained by the oscillation localization model. The red features in the figures increase the probability value of the unit being the oscillation source, while the blue features decrease the probability value. The width represents the SHAP value of the feature. Due to limited display space, only some important features are shown in the figures.

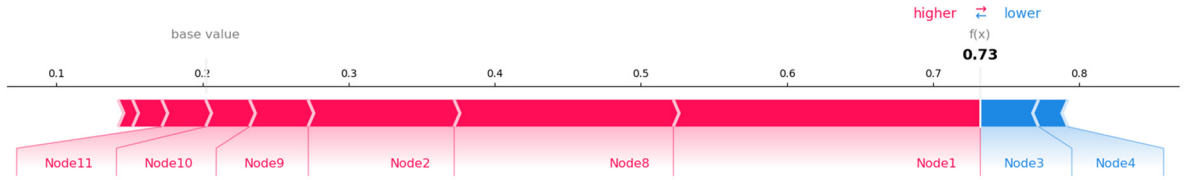


Figure 5. SHAP values of characteristic quantities at each node when G1 is the oscillation source.

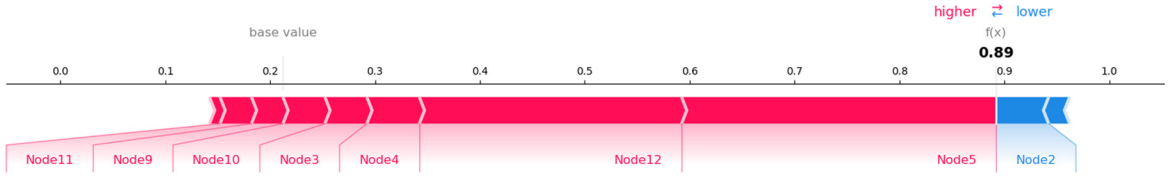


Figure 6. SHAP values of features at each node when G5 is the oscillation source.

(1) In both cases, the baseline SHAP values for each node's features are around 0.2. This reflects the 20% probability of randomly identifying the oscillation source among the five generator units when the model is untrained, validating the SHAP method's effectiveness.
(2) When G1 is the oscillation source, nodes 1, 2, and 8 have the highest SHAP values; when G5 is the source, nodes 4, 5, and 12 are most significant. This shows that, regardless of whether the oscillation unit is a synchronous generator or a wind turbine, nearby nodes' features contribute the most to localization decisions.
(3) Transmission network nodes 9, 10, and 11 have positive but minor contributions. Depending on the situation, these nodes could be considered for removal.

4.4. Improvement of the Wide-Frequency Oscillation Localization Model

According to the global interpretation, the global SHAP values of the current signal are not high, and their contribution to the localization model is relatively small. Therefore, considering the model size and computational costs, the current signal features for all nodes are removed. At the same time, as shown in Figures 5 and 6, although the measurement information at the nodes on the transmission network side has a positive effect on localization accuracy, its impact is relatively small, so they are also removed. Only the nodes at the generator outlets and the nodes connected to the transmission network near the generator units are retained. Thus, the input features of the wide-frequency oscillation localization model are as shown in Table 2.

Table 2. Input Features before and after model improvement.

	Original Model Input Features	Improved Model Input Features
Number of Nodes	Nodes 1–15	Nodes 1–8 and 12–15
Electrical Parameters	Voltage Magnitude, Voltage Phase Angle, Current Magnitude, Current Phase Angle, Active Power, Reactive Power	Voltage Magnitude, Voltage Phase Angle, Active Power, Reactive Power

To validate the effectiveness of the proposed improvements, a comparative analysis is conducted using global feature screening. The models compared are:

Model 1: The original wide-frequency oscillation localization model.

Model 2: Based on Model 1, with only the number of nodes modified as per Table 2.

Model 3: Based on Model 1, with only the electrical parameters of each node modified as per Table 2.

Model 4: Based on Model 1, with all improvements listed in Table 2 applied.

Model 5: Based on Model 4, with an additional feature attention mechanism layer before the LSTM network layer.

All models are trained and tested on the same datasets to assess their performance, with results shown in Table 3.

Table 3. Positioning results before and after model improvement.

Model	Localization Accuracy (%)	Computation Time (s)
Model 1	96.93	213
Model 2	96.71	184
Model 3	97.74	169
Model 4	98.22	152
Model 5	98.91	158

Table 3 shows that Model 1 and Model 2 have similar localization accuracy, indicating that the removed node features have minimal impact on the oscillation localization model. However, reducing input features significantly decreases computation time. From Model 1 to Model 4, feature reduction is progressively applied, effectively lowering data dimensionality, reducing modeling workload, and decreasing the number of neuron parameters, thereby further reducing computation time. By removing weakly correlated features, the localization model focuses entirely on critical features, improving accuracy.

Comparing Model 4 and Model 5, the added feature attention mechanism slightly increases computation time but enhances the model's focus on key features during training, such as active power at generator outlets and voltage magnitude and phase angle at grid connection points. As a result, the improved model's accuracy reaches nearly 99%, demonstrating the effectiveness and practical value of the proposed enhancements for the wide-frequency oscillation localization model.

The four-machine two-area system case study offers a solid proof of concept foundation. To apply the model to large-scale power systems, enhancements are required in model architecture, particularly in optimizing hierarchical aggregation strategies to manage computational complexity from increased node numbers. LSTM's temporal feature extraction is useful for cross-regional oscillation propagation, but it should be combined with spatio-temporal attention mechanisms for better focus on key paths.

Regarding the interpretability and simplification techniques, we believe they will remain effective in more realistic scenarios. The core principles of our method, which focus on identifying key variables and leveraging domain knowledge, are robust and scalable. However, we also recognize that as systems grow in size and complexity, the need for more advanced visualization and diagnostic tools may arise. We plan to explore these enhancements in future work, ensuring that our method remains practical and insightful for a wide range of power system applications.

5. Conclusions

To address the issue of insufficient model interpretability, this paper proposes an interpretability framework for the wide-frequency oscillation localization model based on SHAP values, explaining the model's operation from both global and local perspectives. Based on this, a feature attention mechanism is introduced to enhance the model. Further analysis using the four-machine, two-area simulation system yields the following conclusions:

(1) Global SHAP values indicate that the power and voltage signals at the outlets of synchronous generators and wind turbines significantly influence the model's localization results, contributing highly to the accuracy of wide-frequency oscillation localization. In contrast, the oscillation features contained in the current signal contribute relatively little to localization accuracy.
(2) Local SHAP values show that the features at nodes near the oscillation source contribute the most, while nodes on the transmission network side have little significance for the model's localization effectiveness. Additionally, for different oscillation source units, the measurement data from some nodes have a reverse effect.
(3) After introducing the feature attention mechanism, the model's accuracy is significantly improved, indicating that the model focuses more on key features during training. The improvement strategy for the wide-frequency oscillation localization model is effective and practical.

Author Contributions

Y.L.: conceptualization, methodology, software, writing—reviewing and editing; J.G.: data curation, writing—original draft preparation; J.W.: visualization; Z.J.: supervision, methodology; H.W.: validation. All authors have read and agreed to the published version of the manuscript.

Funding

This research was supported by National Science Foundation of China (No. 52307119).

Data Availability Statement

The simulation data contain confidential information and are not publicly available. A summary of the results is presented in the manuscript, and the simulation framework is described in the Methods section. For access to the data, please contact the corresponding author via email.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Cheng, H.; Li, J.; Wu, Y.; et al. Challenges and Prospects of AC/DC Transmission Network Planning Considering High Penetration of Renewable Energy. *Autom. Electr. Power Syst.* **2017**, *41*, 19–27.
2. Wang, M.; Sun, H. Online Localization Analysis Technology for Forced Power Oscillation Sources. *Proc. CSEE* **2014**, *34*, 6209–6215.
3. Banna, H.U.; Solanki, S.K.; Solanki, J. Data-driven Disturbance Source Identification for Power System Oscillations Using Credibility Search Ensemble Learning. *IET Smart Grid* **2019**, *2*, 293–300.
4. Gu, J.; Xie, D.; Gu, C.; et al. Location of Low-Frequency Oscillation Sources Using Improved D-S Evidence Theory. *Int. J. Electr. Power Energy Syst.* **2021**, *125*, 106444.
5. Doran, D.; Schulz, S.; Besold, T.R. What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. *arXiv* **2017**, arXiv:1710.00794.
6. Samek, W.; Wiegand, T.; Müller, K.R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *arXiv* **2017**, arXiv:1708.08296.
7. Zhang, Q.-S.; Zhu, S.-C. Visual Interpretability for Deep Learning: A Survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39.
8. Han, T.; Chen, J.; Li, Y.; et al. Research on Interpretable Proxy Models for Power System Stability Assessment Using Machine Learning. *Proc. CSEE* **2020**, *40*, 4122–4131.
9. Ji, S.; Li, J.; Du, T.; et al. A Review of Interpretability Methods, Applications, and Security Research for Machine Learning Models. *J. Comput. Res. Dev.* **2019**, *56*, 2071–2096.
10. Saltelli, A.; Tarantola, S.; Campolongo, F.; et al. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. *J. R. Stat. Soc. Ser. A* **2004**, *168*, 464.
11. 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, CA, USA, 13–17 August 2016.
12. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
13. Lundberg, S.M.; Erion, G.G.; Lee, S.I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**,

arXiv:1802.03888.
14. Robnik-Šikonja, M.; Kononenko, I. Explaining Classifications for Individual Instances. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 589–600.
15. Fong, R.; Vedaldi, A. Interpretable Explanations of Black Boxes by Meaningful Perturbation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
16. Li, J.; Monroe, W.; Jurafsky, D. Understanding Neural Networks through Representation Erasure. *arXiv* **2016**, arXiv:1612.08220.
17. Chen, W. Pre-Loan Overdue Identification and Model Expression for Internet Finance Based on XGBoost. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2019.
18. Yu, J. Research on Prediction Model for Gestational Diabetes Based on Ensemble Learning Algorithms. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2019.
19. Feng, S.; Cui, H.; Chen, J.; et al. Wide-Frequency Oscillation Disturbance Source Localization Method Based on Autoencoder Signal Compression and LSTM. *Autom. Electr. Power Syst.* **2022**, 1–12.