Article

Transmission Line Icing Prediction Based on Physically Guided Fast-Slow Transformer

Feng Wang [1] and Ziming Ma [2,*]

[1] College of Civil Engineering and Architecture, China Three Gorges University, Yichang 443002, China; wanggoody@126.com

[2] College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China

* Correspondence: zmma1070@ctgu.edu.cn

Abstract: To improve the accuracy of the icing prediction model for overhead transmission lines, a physics-guided Fast-Slow Transformer icing prediction model for overhead transmission lines is proposed, which is based on the icing prediction model with meteorological input characteristics. First, the ice cover data is segmented into different time resolutions through Fourier transform; a transformer model based on Fourier transform is constructed to capture the local and global correlations of the ice cover data; then, according to the calculation model of the comprehensive load on the conductor and the conductor state equation, the variation law of ice thickness, temperature, wind speed, and tension is analyzed, and the model loss function is constructed according to the variation law to guide the training process of the model. Finally, the sample mixing enhancement algorithm is used to reduce the overfitting problem and improve the generalization performance of the prediction model. The results show that the proposed prediction model can consider the mechanical constraints in the ice growth process and accurately capture the dependence between ice cover and meteorology. Compared with traditional prediction models such as LSTM (Long Short-Term Memory) networks, its mean square error, mean absolute error, and mean absolute percentage error are reduced by 0.464–0.674, 0.41–0.53, and 8.87–11.5%, respectively, while the coefficient of determination (R^2) is increased by 0.2–0.29.

Keywords: icing prediction; Fourier transform; attention mechanism; physical guidance; Mixup

1. Introduction

As the "artery" of the power system, the safe operation of overhead transmission lines is of great significance. However, China's terrain is complex and diverse. In the natural environment, transmission lines are prone to icing disasters, threatening the safe operation of transmission lines. For example, in 2023, the power transmission lines in the North China country were severely iced, causing 4 high-voltage transmission lines to fail and more than 100,000 households to lose power.

At present, domestic and foreign scholars have conducted extensive research on the prediction of icing on overhead transmission lines. Ice prediction models are mainly divided into models based on physical mechanisms, models based on statistical methods, and models based on neural networks. Based on thermodynamics and fluid mechanics, some scholars have studied the physical process of ice formation on transmission lines, explored its physical mechanism, and established prediction models based on physical mechanisms, such as the Goodwin model [1] and the Makkonen model [2], etc. However, due to the complexity of the wire icing process, the overhead transmission line icing

prediction model based on physical mechanisms has been simplified to a certain extent, and important parameters in the model, such as droplet size, are not correct in the actual process as they are difficult to measure, which makes such models less accurate. Some scholars have established an icing prediction model based on statistical methods by analyzing the linear relationship between icing time series and meteorological time series. Sun Wei et al. [3] used wavelet transform to denoise the original ice data and used an extreme learning machine optimized by the bat algorithm to predict the ice thickness, and Chen Yong et al. [4] used principal component analysis to extract effective information from meteorological data and established an ice prediction model based on LSSVM. However, icing prediction models based on statistical methods often require data distribution and variable distribution to meet certain assumptions and have limited prediction capabilities for nonlinear and high-dimensional complex data. They cannot adapt to modeling complex ice processes and have insufficient prediction accuracy.

In recent years, the use of neural networks to predict the icing of overhead transmission lines has attracted increasing attention. Wang Xunting et al. [5], Li Xianchu et al. [6], and D Niu et al. [7] established an ice prediction model based on the BP neural network. They used the powerful nonlinear mapping ability of the BP neural network to learn the nonlinear mapping relationship between ice and meteorological data. Li Bo et al. [8], L Li et al. [9], and Chen Lifan et al. [10] established an ice prediction model based on the convolutional neural network. By stacking convolutional layers and pooling layers, they captured the local correlation of ice data in time and space, learned multi-scale feature representation, and thus achieved accurate modeling of ice and meteorological data. Su Renbin et al. [11], Chen Bin et al. [12], and Yu Tong et al. [13] established an ice prediction model based on the recurrent neural network, predicting future ice risks by capturing the time correlation between variables. In addition, some scholars combined the physical laws of the ice process with neural network models. For example, Yu Tong et al. [14] conducted a force analysis on the transmission line, established a comprehensive load calculation model for the line, analyzed the changing laws of ice thickness, wind deflection angle, and comprehensive load, and constructed a model loss function based on the changing laws to guide the model training process. Wang F et al. [15] tried to introduce physical laws into the prediction process of GRU to avoid conflicts between the prediction results and the changing laws between the tension and ice thickness of the transmission line. However, the above prediction models have limited modeling capabilities and cannot adapt to the complex icing process. It is difficult to accurately capture the correlation between icing thickness and meteorological factors. The prediction results do not match the actual icing law.

Therefore, this paper proposes a physically guided FSFormer (Fast-Slow Transformer) overhead transmission line icing prediction model. First, the features are segmented according to different scales through the Fourier transform, and FSFormer is used to capture local correlation and overall correlation, respectively; then, the changing law of icing thickness, temperature, wind speed, and tension of transmission lines is studied, and the loss function is constructed based on this law. In the model training process, the sample mixed data enhancement algorithm (Mixup) is combined. The research results can provide guidance for the power sector in formulating anti-icing and de-icing measures.

2. Materials and Methods

2.1. Ice Cover Prediction Model Based on Physical Laws

For traditional icing prediction models, meteorological factors are generally used as features to predict wire icing thickness. However, some physical parameter data such as tension that can be collected by the transmission line monitoring system are not effectively

used. Therefore, this article attempts to combine the transmission line mechanics model to analyze the change rules between the conductor ice thickness, temperature, wind speed, and tension, and combine the change rules with the model to play a correction role in the ice prediction process so that the model prediction process is more accurate. This complies with the ice formation process and avoids prediction results that violate the rules.

2.1.1. Static Model of Wind and Ice Loads on Transmission Lines

First, the stress analysis of the overhead transmission line is carried out. It is usually subjected to three kinds of loads in the outside world, namely the deadweight of the conductor, the weight of ice caused by icing, and the lateral load caused by wind pressure.

In Figure 1: d is the conductor diameter, b is the ice thickness, γ_3 is the vertical-specific load when the ice thickness is b, γ_5 is the horizontal-specific load when the ice thickness is b and the wind speed is v, and γ_7 is the comprehensive specific load when the ice thickness is b and the wind speed is v.
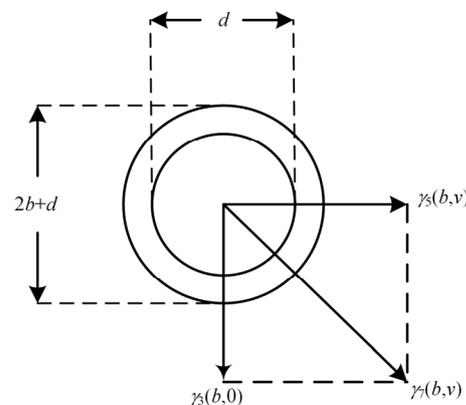

Figure 1. Transmission line stress model.

1.   Vertical comprehensive load ratio of ice-covered conductor

For the comprehensive load ratio of the conductor under ice conditions, to simplify the calculation, the shape of the ice coating is approximately considered to be circular, and its calculation equation is:

$$\gamma_3(b,0) = 27.728\frac{b(b+d)}{A} \times 10^{-3} + \frac{qg}{A} \times 10^{-3} \tag{1}$$

where q is the mass of the conductor per unit length; g is the acceleration of gravity; and A is the cross-sectional area of the overhead line.

2.   Wind pressure load ratio during ice cover

The specific load caused by wind pressure during ice cover is calculated as follows:

$$\gamma_5(b,v) = \beta_c\alpha_f\mu_{sc}B(d+2b)\frac{0.625v^2}{A}\sin^2\theta \times 10^{-3} \tag{2}$$

where β_c is the wind load adjustment coefficient; α_f is the wind speed unevenness coefficient; μ_{sc} is the wind load shape coefficient; v is the wind speed; θ is the angle between the wind direction and the conductor; and B is the ice-wind load enhancement coefficient.

3.   Comprehensive load ratio

The calculation formula for combining the vertical-specific load and the horizontal-specific load is:

$$\gamma_7(b,v) = \sqrt{\gamma_3^2(b,0) + \gamma_5^2(b,v)} \tag{3}$$

Assuming that the transmission line is an ideal flexible line, the load on the overhead line is evenly distributed, and the overhead line is a completely elastic body with its elastic coefficient remaining unchanged, the state equation of the conductor can be obtained as follows:

$$\sigma_2 - \frac{E\gamma_2^2 l^2 \cos^3 \beta}{24\sigma_2^2} = \sigma_1 - \frac{E\gamma_1^2 l^2 \cos^3 \beta}{24\sigma_1^2} - \alpha E \cos \beta (t_2 - t_1) \tag{4}$$

where σ is the horizontal stress of the transmission line; E is the elastic modulus; γ_2 is the specific load of state 2; l is the span; β is the horizontal angle of the span; and α is the linear expansion coefficient.

Equation (4) can be used to obtain the variation patterns of ice thickness, temperature, wind speed, and tension of transmission lines under different meteorological conditions.

2.1.2. Correction Method for Physical Guidance

According to the above rules, the model is corrected during the training process. The correction is implemented in the form of a comprehensive loss function. The model's current predicted value of ice thickness, wind speed, temperature, and tension monitoring data and the actual ice thickness, wind speed, temperature, and tension data at the previous moment are substituted into Equation (4). The absolute value of the calculation result is defined as the degree of violation of the physical law. This is used to correct the model's training process to make it more consistent with the actual conductor icing process. At the same time, considering that some factors are ignored in the analysis process, a certain threshold value is set for the degree of violation of the physical law to increase the stability of the training process.

Assume that the loss function equation during model training is:

$$loss_{\text{train}} = loss_{\text{model}} + \alpha loss_{\text{phy}} \tag{5}$$

where $loss_{\text{train}}$ is the comprehensive loss function; $loss_{\text{model}}$ is the model loss function, which indicates the closeness between the predicted value and the actual value; $loss_{\text{phy}}$ is the physical law loss function; and α indicates the weighting of the physical law loss function.

The model loss function uses the mean square error, and its equation is:

$$loss_{\text{model}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{6}$$

In the process of model training, in order to make the prediction process of the prediction model and the actual conductor ice growth process as consistent as possible, the predicted ice thickness, temperature, wind speed, tension at the current moment, and the actual ice thickness, temperature, wind speed, tension at the previous moment are substituted into the conductor state equation to obtain the loss function that introduces physical laws. The calculation equation is:

$$loss_{\text{phy}} = \left[\sigma_2 - \frac{E\gamma_2^2 l^2 \cos^3 \beta}{24\sigma_2^2} - \sigma_1 + \frac{E\gamma_1^2 l^2 \cos^3 \beta}{24\sigma_1^2} + \alpha E \cos \beta (t_2 - t_1) \right]^2 \tag{7}$$

$$loss_{\text{phy}} = \max(loss_{\text{phy}}, a) \tag{8}$$

where a is the set positive threshold value.

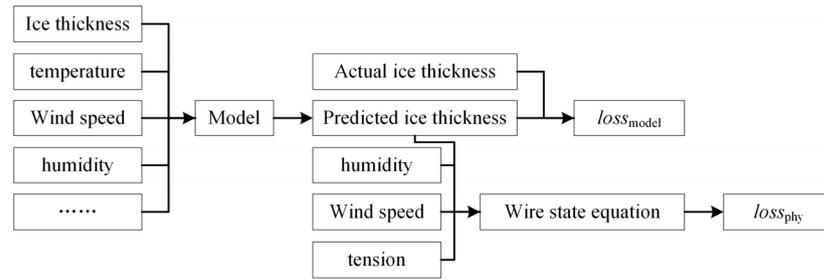The structure of the comprehensive loss function is shown in Figure 2.

Figure 2. Loss function structure diagram.

2.2. Prediction Model Structure

The overall structure of the prediction model proposed in this paper is shown in Figure 3, which includes the Mixup data enhancement module, FSFormer, and a loss function containing physical laws. The Mixup data enhancement module is used to enrich the training samples, then the correlation between ice cover and meteorological factors is extracted through FSFormer, and the model prediction results are corrected by introducing a loss function containing physical laws.
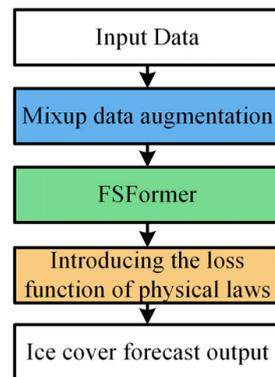

Figure 3. Prediction flowchart.

2.3. FSFormer

The Transformer model is widely used in the field of time series prediction. It relies on the self-attention mechanism to calculate input and output. The network structure of the Transformer model is divided into an encoder and a decoder. In order to identify the position information of the time series, position encoding is generally added to the input. More and more studies have shown that using an encoder to extract features and using a fully connected layer as a decoder can achieve the same effect as the original Transformer. This article uses an encoder to extract the correlation between meteorological data and ice thickness. We use a fully connected layer to predict ice thickness based on the extracted features. Its overall structure is shown in Figure 4.

2.3.1. Adaptive Segmentation Based on Fourier Transform

Fourier transform is a common signal processing method that can convert signals from the time domain to the frequency domain. The Fourier transform equation is:

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{9}$$

where $f(t)$ represents the original time signal and $F(j\omega)$ represents the frequency domain representation of the original time signal.

For discrete signals that cannot be continuously integrated, the discrete Fourier transform is often used. The formula for the discrete Fourier transform is:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}nk} \tag{10}$$

where $x[n]$ represents the original discrete time signal; N represents the number of original time signals; and k represents the signal period.
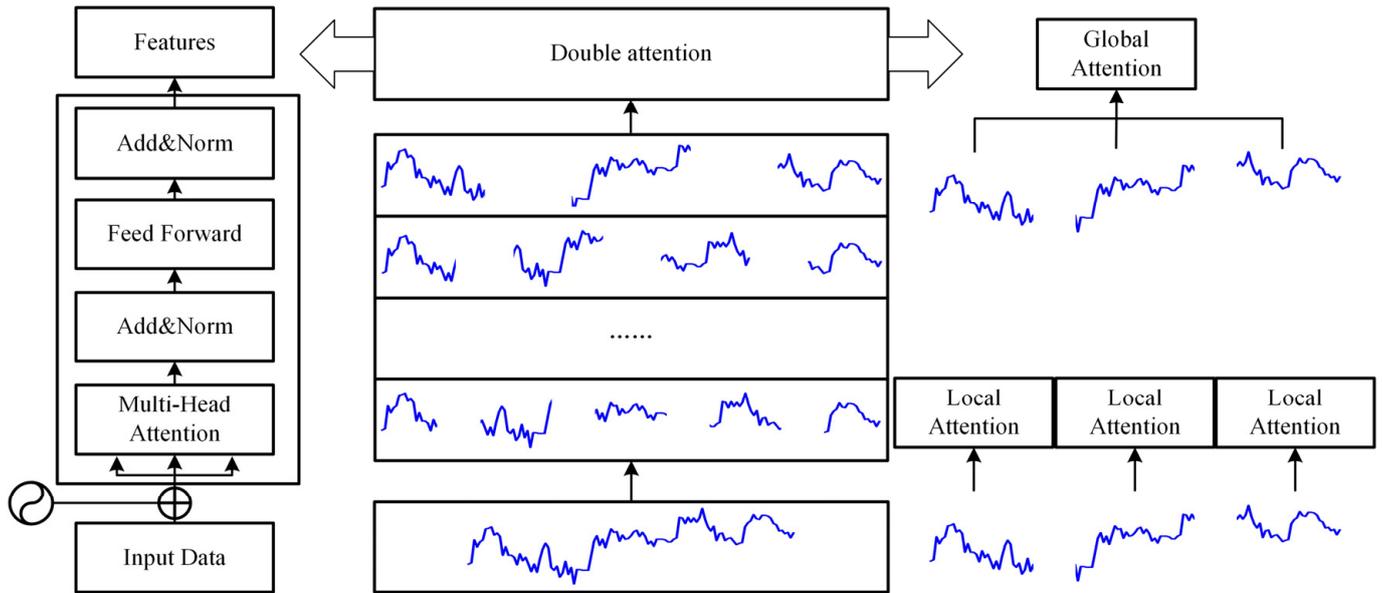

Figure 4. FSFormer structure.

For a time series X, whose shape is $T \times C$, the frequency domain representation of X is obtained by the Fast Fourier Transform, and the result is averaged to obtain the frequency domain representation of X [16]. The calculation formula is:

$$\begin{bmatrix} y_1 & \cdots & y_T \end{bmatrix} = \text{Avg}(\text{FFT}(X)) \tag{11}$$

where X represents the historical ice cover data; FFT represents the Fast Fourier Transform; and Avg represents the average function.

Take the first k values with the largest modulus value to obtain the representative period of the historical ice cover data. According to the representative period P, the ice cover data is segmented. The original ice coverage data is converted into the form of $B \times C \times P \times N$, where P is the period length and N is the number of segments.

2.3.2. Dual Attention Mechanism

For the converted time series, a dual attention mechanism is adopted. The local attention module is used to extract the local correlation within the segment, and the global attention module is used to extract the global correlation of the entire sequence. The transmission line ice data is decomposed into subsequences of different fragment sizes through the Fourier transform. Under the guidance of fragment division, the time dependency is modeled from different scales, thereby achieving an accurate capture of the correlation of ice data [17].

As shown in Figure 4, for local attention, based on each segment of shape $C \times P$, the trainable query matrix Q, key-value matrix K, and value matrix V are used to calculate the local attention within the segment. The calculation process is as follows:

$$\text{Atten}_i^{\text{local}} = \text{Sotfmax}\left(\frac{Q_i^{\text{local}}\left(K_i^{\text{local}}\right)^T}{\sqrt{d_m}}\right) V_i^{\text{local}} \tag{12}$$

where Q, K, V are the query matrix, key-value matrix, and value matrix within the segment, and d_m is the vector dimension.

For the global attention between segments, each segment is first flattened into a vector and then restored to its original shape after calculating the global attention. The calculation process is as follows:

$$\text{Atten}^{\text{global}} = \text{Sotfmax}\left(\frac{Q^{\text{global}}\left(K^{\text{global}}\right)^T}{\sqrt{d'_m}}\right) V^{\text{global}} \tag{13}$$

where Q, K, V are the global query matrix, key-value matrix, and value matrix, and d'_m is the vector dimension.

Then the local attention result is added to the global attention result to get the final attention result.

2.4. Mixup Data Augmentation

The performance of the data-driven neural network prediction model is affected by the quality of the training samples. However, due to the limited observation cost, it is difficult to obtain relevant data on the icing of overhead transmission lines. This results in the icing prediction model being prone to overfitting and poor generalization performance. Therefore, this paper adopts the data enhancement Mixup algorithm [18] to improve the generalization performance and robustness of the prediction model.

Data enhancement is commonly used in the field of computer vision [19,20]. In recent years, Mixup has received increasing attention in the field of time series prediction [21,22]. The Mixup data enhancement method is based on the principle of neighborhood risk minimization and generates virtual data by linearly interpolating the original data. The Mixup algorithm improves the model's ability to predict unknown samples. The specific method is as follows:

$$\widetilde{x} = \lambda x_i + (1 - \lambda)x_j \tag{14}$$

$$\widetilde{y} = \lambda y_i + (1 - \lambda)y_j \tag{15}$$

where $\lambda \in [0, 1]$, $\widetilde{x}$, and $\widetilde{y}$ are the generated virtual samples and labels; $\lambda \in \text{Beta}(\alpha, \alpha)$; and α controls the degree of linear interpolation.

2.5. Experimental Data and Experimental Settings

The experimental data used in this paper comes from the online monitoring system for icing transmission lines. The monitoring time is from 17 January to 25 January 2012, and the monitoring data collection frequency is once every 20 min, including a complete ice coating process. Some data are shown in Table 1. There are 539 sets of monitoring data. Among them, the minimum ice thickness is 0 mm, the minimum temperature is $-16.12\,^{\circ}\text{C}$, the minimum humidity is 60.07%, the minimum wind speed is 0 m/s, the minimum light intensity is 60.61 Lux, and the minimum pressure is 66.46 kPa. The maximum ice thickness is 30 mm, the maximum temperature is 2.43 $^{\circ}$C, the maximum humidity is 93.96%, the maximum wind speed is 14.53 m/s, the maximum light intensity is 162.51 Lux, and

the maximum pressure is 66.50 kPa. The average ice thickness is 11.17 mm, the average temperature is -7.22 °C, the average humidity is 74.45%, the average wind speed is 3.09 m/s, the average light intensity is 70.35 Lux, and the average pressure is 66.48 kPa. The standard deviations of ice thickness, temperature, humidity, wind speed, light intensity, and pressure are 9.83 mm, 4.10 °C, 7.37%, 2.67 m/s, 20.70 Lux, and 0.01 kPa, respectively.

Table 1. Ice monitoring data.

Serial Number	Icing Thickness (mm)	Humidity (%)	Temperature (°C)	Wind Speed (ms^{-1})	Illumination (Lux)	Air Pressure (kPa)	Tension (N)
1	0.03	62.86	0.17	4.54	63.79	66.47	28,621
2	0.39	62.11	0.12	5.51	62.04	66.47	29,094
3	0.22	61.99	0.07	5.93	64.90	66.47	28,888
...	...	...	...	...	...	...	
537	4.85	65.95	2.43	1.94	62.97	66.47	35,471
538	4.75	65.96	2.33	0.63	61.87	66.46	35,307
539	4.65	66.93	2.30	3.03	62.03	66.46	35,149

In this paper, the time step is set to 10, the step feature is set to 6, including ice thickness, temperature, humidity, wind speed, light, and pressure, and the output sequence length is set to 1. The first 60% of the data is used as the training set, the second 20% of the data is used as the validation set, and the third 20% of the data is used as the test set. Obviously, abnormal data were removed, missing data were interpolated and supplemented, and the maximum normalization method was used to unify the dimensions of each feature.

This paper uses mean square error (*MSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), and coefficient of determination (R^2) as evaluation indicators of experimental results. The calculation equation is:

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{17}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2} \tag{19}$$

where $\hat{y}i_i$ is the model prediction value; y_i is the actual ice thickness; N is the number of samples; and $\overline{y}_i$ is the mean icing thickness.

All models are implemented in Python 3.7, with epoch set to 300, batch size set to 32, and learning rate set to 0.002. During the training process, the model in this article uses a comprehensive loss that introduces the physical laws function, while other models use the mean square error as the loss function. We use the Adam optimizer to update the parameters of the model. Prediction models such as BP neural network, RNN, LSTM, TCN, CNN-LSTM, Attention-BiLSTM, and CNN-BiGRU were selected for comparison. The main parameter settings of each model are shown in Table 2.

Table 2. Prediction model parameter table.

Prediction Model	Main Parameters of the Model
Proposed model	$\alpha = 0.5$, $k = 3$, n_layers = 3, alpha = 5, hidden_size = 256, num_heads = 8
BP	hidden_size = 256
RNN	n_layers = 3 hidden_size = 256
LSTM	n_layers = 3 hidden_size = 256
TCN	n_channels = 32, n_layers = 3
CNN-LSTM	hidden_size = 256, n_layers = 3, kernel_size = 2
Attention-BiLSTM	n_layers = 3, hidden_size = 256, num_heads = 8, n_layers = 3
CNN-BiGRU	hidden_size = 256, n_layers = 3, kernel_size = 2

3. Results

### 3.1. Performance Comparison of Traditional Prediction Models

Each transmission line icing prediction model is trained on the data set, and the prediction error indicators of the model are shown in Table 3.

Table 3. Model evaluation indicators.

Model	MSE (mm^2)	MAE (mm)	MAPE	R^2
Proposed model	0.096	0.24	4.87%	0.96
BP	0.75	0.69	14.74%	0.68
RNN	0.77	0.77	16.37%	0.67
LSTM	0.71	0.71	15.03%	0.70
TCN	0.76	0.72	15.04%	0.67
CNN-LSTM	0.66	0.67	14.07%	0.72
Attention-BiLSTM	0.56	0.65	13.74%	0.76
CNN-BiGRU	0.63	0.65	14.34%	0.73

As shown in Table 3, compared with the traditional ice cover prediction model, the model in this paper has the smallest prediction error and the best model performance. Compared with other models, its MSE is reduced by 0.464–0.674, MAE is reduced by 0.41–0.53, MAPE is reduced by 8.87–11.5%, and R^2 is improved by 0.2–0.29.

In order to show the prediction results more clearly, the comparison between the prediction results of each model and the true value on the test set is shown in Figure 5, and Figure 6 is the error box plot of each model result.
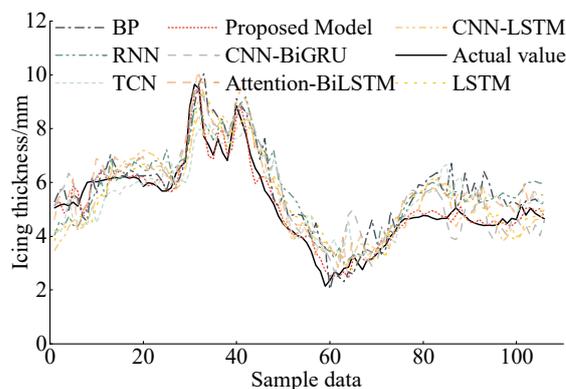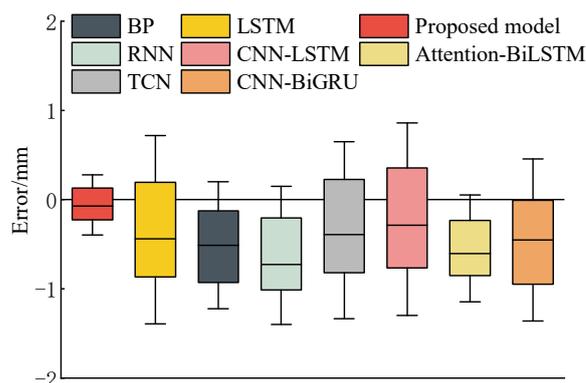

Figure 5. Comparison of prediction results.

Figure 6. Prediction error box plot.

As shown in Figure 5, compared with other models, the prediction model proposed in this paper can better fit the real ice thickness data, especially in the interval where the ice thickness changes more dramatically. The reason is that the Fourier transform divides the input data into segments, which helps FSFormer capture the local correlation within the segments and the global correlation between the segments respectively; secondly, starting from the state equation of the transmission line, this paper introduces physical law constraints in the model loss function, which makes the prediction process of the model closer to the real ice change process. At the same time, the Mixup algorithm also enriches the distribution space of the training data, which helps to achieve accurate prediction of ice on the transmission line under data scarcity.

As shown in Figure 6, the box corresponding to the prediction error of the model proposed in this paper is flatter, and the error median is closer to 0, indicating that the volatility of its prediction error is the smallest. The error distribution medians of the model proposed in this paper are close to 0, indicating that the error stability of this model is better. The median errors of other models are far from 0, indicating that they are less capable of capturing the correlation of ice cover processes and have poor prediction stability.

3.2. Ablation Experiment

In order to verify the effectiveness of each module of the prediction model proposed in this paper, an ablation experiment was conducted on the original data set, and the parameters of the ablation experiment were consistent with the previous article.

The first ablation experiment conditions are shown in Table 4, which are, respectively, removing the loss function introduced with physical laws, FSFormer and Mixup modules (De-FS&Mix&Phy), removing FSFormer and Mixup modules (De-FS&Mix), removing Mixup modules (De-Mix), and the model proposed in this paper. The inputs of each model are kept consistent. The prediction results of each model are shown in Table 4.

Table 4. Ablation experiment 1 indicators.

Working Conditions	Model	MSE (mm^2)	MAE (mm)	MAPE	R^2
1	Proposed model	0.096	0.24	4.87%	0.96
2	De-Mix	0.21	0.36	7.67%	0.91
3	De-FS&Mix	0.48	0.53	11.49%	0.80
4	De-FS&Mix&Phy	0.75	0.69	14.74%	0.68

As shown in Table 4, the model in this paper achieved the best results. After removing the Mixup module, the MSE, MAE, MAPE, and R^2 of the model decreased by 0.114, 0.12, 2.8%, and 0.05, respectively, indicating that the Mixup algorithm generates new samples by mixing samples, which expands the distribution space of the training set. For the ice

prediction model with scarce training data, it significantly improves the accuracy and generalization performance of the model; after removing the FSFormer module, the MSE, MAE, MAPE, and R^2 of the model decreased by 0.27, 0.17, 3.82%, and 0.11, respectively, indicating that the use of the Fourier transform to segment the input features helps the model extract the correlation of the sequence and effectively improves the accuracy of wire ice prediction; after removing the loss function that introduces physical laws, the MSE, MAE, MAPE, and R^2 of the model decreased by 0.27, 0.16, 3.25% and 0.12, respectively, indicating that the introduction of the ice change process constraints derived from the state equation in the loss function helps the model learn the correlation between ice and meteorology, making its prediction process more in line with the actual ice process.

In order to verify the effectiveness of the FSFormer proposed in this paper, the second ablation experiment was conducted, and five working conditions were set. The calculation results are shown in Table 5. Among them, FSFormer_2, FSFormer_3, FSFormer_4, and FSFormer_5 represent the prediction models with segment lengths artificially set to 2, 3, 4, and 5.

Table 5. Ablation experiment 2 indicators.

Working Conditions	Model	MSE (mm^2)	MAE (mm)	MAPE	R^2
1	Proposed model	0.096	0.24	4.87%	0.96
2	FSFormer_2	0.56	0.56	11.28%	0.76
3	FSFormer_3	0.49	0.58	13.13%	0.79
4	FSFormer_4	0.57	0.60	11.79%	0.75
5	FSFormer_5	0.58	0.59	13.34%	0.75

As shown in Table 5, when the length of the segment is artificially set to 2, 3, 4, and 5, the MSE of the model decreases by 0.464, 0.394, 0.474, and 0.484, respectively, which shows the effectiveness of using Fourier transform to adaptively segment the input data. The strategy of manually selecting the segment length will significantly reduce the accuracy of model prediction; at the same time, the traditional direct extraction of correlation from the original sequence is converted to Fourier-assisted segmentation to extract local correlation and global correlation respectively, which improves the model's ability to capture time dynamics.

In order to further illustrate the effectiveness of the physical law loss function proposed in this paper, the physical law loss function is added and removed in this model and the traditional model, respectively, and MSE and $loss_{phy}$ are used as evaluation indicators. The experimental results are shown in Table 6 and Figure 7.

Table 6. Ablation experiment 3 indicators.

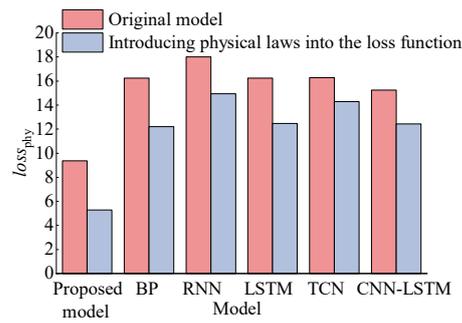| Model	MSE (mm^2)	MAE (mm)	MAPE	R^2
Proposed model	0.27	0.43	8.99%	0.88
PG_Proposed model	0.096	0.24	4.87%	0.96
BP	0.75	0.69	14.74%	0.68
PG_BP	0.48	0.53	11.49%	0.80
RNN	0.77	0.77	16.37%	0.67
PG_RNN	0.57	0.66	13.94%	0.75
LSTM	0.71	0.71	15.03%	0.70
PG_LSTM	0.48	0.56	11.64%	0.79
TCN	0.76	0.72	15.04%	0.67
PG_TCN	0.57	0.62	13.02%	0.76
CNN-LSTM	0.66	0.67	14.07%	0.72
PG_CNN-LSTM	0.51	0.56	12.13%	0.78

Figure 7. Model physical inconsistency.

It can be seen from Table 6 that, compared with the original model, the MSE of our model, BP, RNN, LSTM, TCN, and CNN-LSTM introduced into the loss function are reduced by 0.174, 0.27, 0.2, 0.23, 0.19, and 0.15, MAE is reduced by 0.19, 0.16, 0.11, 0.15, 0.1, and 0.11, MAPE is reduced by 4.12%, 3.25%, 2.43%, 3.39%, 2.02%, and 1.94%, respectively, and R^2 increased by 0.08, 0.12, 0.08, 0.09, 0.09, and 0.06, respectively. This shows that the introduction of physical laws into the loss function of the model is not only applicable to the model proposed in this paper but also to traditional models such as BP and LSTM. As shown in Figure 7, the $loss_{phy}$ of each prediction model is reduced by introducing physical laws into the loss function. Introducing physical laws into the loss function makes the prediction model closer to the actual ice growth process, thereby improving the accuracy and authenticity of ice prediction.

3.3. Model Parameter Sensitivity

The amount of historical ice cover data input into the model has a significant impact on the performance of the model. If the data is too little, the model cannot capture trends and changes, and the prediction results are unstable; if the data is too long, too much noise and irrelevant information will be introduced, which may lead to overfitting.

The Fourier component number k value in the prediction model determines the number of components retained after the Fourier transform. If the k value is too small, the high-frequency information and complex features of the ice data will be ignored; if the k value is too large, unnecessary details and noise of the ice data will be retained, which is not conducive to the prediction of the model. In order to select the optimal historical ice data length T and the value of the Fourier component number k, MSE is used as the evaluation index. The results of the comparative experiment are shown in Figure 8.
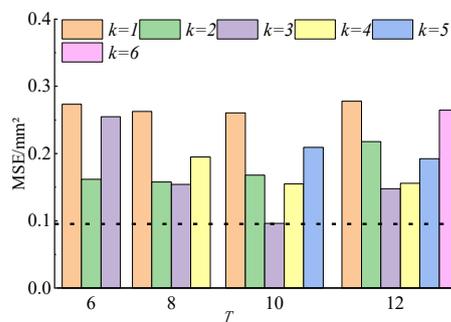

Figure 8. Comparison of performance between different T and k.

As shown in Figure 8, among the historical ice data lengths of 6, 8, 10, and 12, the model with 10-time steps of historical ice data as input has the best performance; for historical ice data inputs of different lengths, as k increases, the MSE of the prediction

model first decreases and then increases; the best prediction model is when T is 10 and k is 3.

The probability density of Beta distribution in the Mixup algorithm is shown in Figure 9. As shown in Figure 9, different values of enhancement parameters result in different probability densities of Beta distribution and different effects of the Mixup enhancement algorithm.
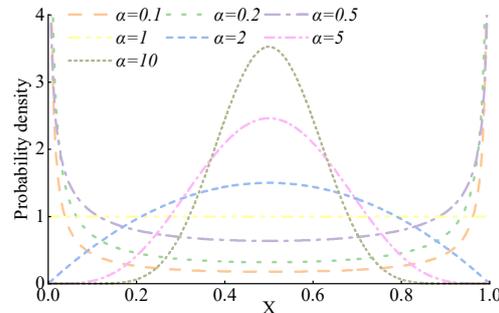

Figure 9. Probability density of Beta distribution at different α.

Using MSE as an indicator, the error of conductor ice coverage prediction when different parameters are added is calculated, and the results are shown in Figure 10.
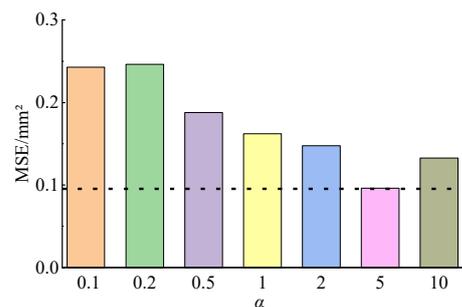

Figure 10. Comparison of different α performances.

As α increases, the MSE of the prediction model shows a trend of first decreasing and then increasing. When α is set to 5, the prediction performance of the model is the best.

4. Discussion

The above experimental results show that the prediction model constructed in this study effectively improves the accuracy and stability of ice thickness prediction. Compared with the traditional prediction model, MSE is reduced by 0.464–0.674, MAE is reduced by 0.41–0.53, MAPE is reduced by 8.87–11.5%, and R^2 is increased by 0.2–0.29. The main reasons are as follows:

- A transmission line stress model is established, and the law of ice change is analyzed according to the conductor state equation. By introducing the physical law constraint into the loss function, the ice prediction process is more in line with the actual ice growth process, which improves the accuracy and authenticity of transmission line ice prediction.
- In view of the complex line icing process, the input historical data is segmented through the Fourier transform, local attention is used to capture local correlation, and global attention is used to capture global correlation. Compared with the traditional model that directly models the input data, the complex problem is decomposed and the accuracy of the icing prediction model is improved.

- Taking into account the difficulty in collecting ice cover monitoring data and the insufficient data for model training, the Mixup data enhancement algorithm is used to expand the distribution space of training data and improve the generalization performance of the model.

The icing process of power transmission lines involves multiple fields such as thermodynamics and fluid mechanics. The process is complex and difficult to model. However, this paper introduces the conductor state equation into the icing prediction model. Through the mechanical constraints of the conductor itself, the prediction process of the icing prediction model is more in line with the actual icing growth process. At the same time, experiments have proved that this method also has a certain effect on the traditional model. However, the physical laws introduced in this paper are simplified and can only consider the constraints of four factors: temperature, wind speed, ice thickness, and tension. They are not applicable to more meteorological factors, such as rain in the actual ice-covering process, and can only consider the ideal circle for the shape of ice. Therefore, in the future, we will consider establishing a more accurate conductor state equation based on more accurate monitoring data, combining more data and features, further enhancing the applicability of physical laws, and providing more accurate support for the safe operation of transmission lines.

5. Conclusions

Aiming at the problem of icing prediction for overhead transmission lines, this paper proposes a prediction model of FSFormer based on physics guidance. The model introduces the conductor mechanical state equation in model training to constrain the prediction process of the icing prediction model, which contributes to the field of disaster prevention and mitigation of transmission lines.

Author Contributions: Conceptualization, F.W. and Z.M.; methodology, F.W. and Z.M.; software, Z.M.; formal analysis, F.W.; writing—original draft preparation, Z.M.; writing—review and editing, F.W.; visualization, Z.M.; supervision, F.W. and Z.M.; funding acquisition, F.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goodwin, E.J.; Mozer, J.D.; DiGioia, A.M.; Power, B.A. Predicting ice and snow loads for transmission line design. In *Proceedings of the International Workshop on Atmospheric Icing of Structures*; Defense Technical Information Center: Fort Belvoir, VA, USA, 1983.
2. Makkonen, L. Models for the growth of rime, glaze, icicles and wet snow on structures. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **2000**, *358*, 2913–2939. [CrossRef]
3. Sun, W.; Wang, C. Staged icing forecasting of power transmission lines based on icing cycle and improved extreme learning machine. *J. Clean. Prod.* **2019**, *208*, 1384–1392. [CrossRef]
4. Chen, Y.; Li, P.; Zhang, Z.; Nie, H.; Shen, X. Online prediction model for power transmission line icing load based on PCA-GA-LSSVM. *Power Syst. Prot. Control.* **2019**, *47*, 110–119.
5. Wang, X.; Ding, J.; Zhang, F.; Sun, H. Research on Transmission Line Icing Prediction Technology Based on Improved BP Neural Network. *Mach. Des. Manuf.* **2024**, *09*, 306–310.
6. Li, X.; Zhang, X.; Liu, J.; Hu, J. Prediction of Transmission Line Icing Thickness Applying AMPSO-BP Neural Network Model. *Electr. Power Constr.* **2021**, *42*, 140–146.

7. Niu, D.; Liang, Y.; Wang, H.; Wang, M.; Hong, W.-C. Icing Forecasting of Transmission Lines with a Modified Back Propagation Neural Network-Support Vector Machine-Extreme Learning Machine with Kernel (BPNN-SVM-KELM) Based on the Variance-Covariance Weight Determination Method. *Energies* **2017**, *10*, 1196. [CrossRef]

8. Li, B.; Li, P.; Gao, L.; Yang, J.; Bao, H. Prediction model for weight of ice coating on transmission line based on PCA-VMD-CNN. *J. Saf. Sci. Technol.* **2022**, *18*, 216–222.

9. Li, L.; Luo, D.; Yao, W. Analysis of transmission line icing prediction based on CNN and data mining technology. *Soft Comput.* **2022**, *26*, 7865–7870. [CrossRef]

10. Chen, L.; Zhang, L.; Song, H.; Chen, K.; Wu, B.; Chen, S.; Sheng, G. Natural Disaster Accident Prediction of Transmission Line Based on Graph Convolution Network. *Power Syst. Technol.* **2023**, *47*, 2549–2557.

11. Su, R.; Xiong, W.; Liu, X.; Zhang, L.; Yu, M.; Zhou, Q.; Cao, M. Study on BP Neural Network Based on a New Metaheuristic Optimization Algorithm and Prediction of Mechanical Response for 500 kV UHV Transmission Lines Considering Icing. *J. Basic Sci. Eng.* **2024**, *32*, 100–122.

12. Chen, B.; Xu, Z.; Jia, Y.; Ding, R.; Zhang, S.; Li, B.; Wang, J. Ice-Cover Prediction Model of Overhead Transmission Conductor Based on VMD-SSA-LSTM. *J. China Three Gorges Univ. (Nat. Sci.)* **2024**, *46*, 105–112.

13. Yu, T.; Li, Y. Prediction Model of Equivalent Ice Thickness of Transmission Line Based on Attention-WOA-BiLSTM. *Technol. Discuss.* **2023**, *01*, 48–54.

14. Yu, T.; Li, Y. Prediction Model Used Physics Guided SSA-BiGRU for Icing Thickness of Transmission Lines. *Electr. Power Sci. Eng.* **2022**, *38*, 28–36.

15. Wang, F.; Lin, H.; Ma, Z. Transmission Line Icing Prediction Based on Dynamic Time Warping and Conductor Operating Parameters. *Energies* **2024**, *17*, 945. [CrossRef]

16. Wang, S.; Wang, S.; Zhao, Q.; Dong, Y. Distributed Wind Power Forecasting Method Based on Frequency Domain Decomposition and Precision-weighted Ensemble. *Electr. Power Constr.* **2023**, *44*, 84–93.

17. Shi, Z.; Ran, Q.; Xu, F. Short-term Load Forecasting Based on Aggregated Secondary Decomposition and Informer. *Power Syst. Technol.* **2024**, *48*, 2574–2583.

18. Lu, Y.; Wang, G.; Huang, S. A short-term load forecasting model based on mixup and transfer learning. *Electr. Power Syst. Res.* **2022**, *207*, 107837–107845. [CrossRef]

19. Xu, M.; Li, H. Face5 series face detection algorithm based on improved YOLOv5s-face. *J. Chongqing Univ. Technol. (Nat. Sci.)* **2024**, *38*, 194–202.

20. Wang, K.; Lou, S.; Wang, Y. Small object detection algorithm based on improved YOLOv3. *J. Appl. Opt.* **2024**, *45*, 732–740.

21. Wang, Y.; Zhang, H. A feature transfer model with Mixup and contrastive loss in domain generalization. *J. Univ. Sci. Technol. China* **2024**, *54*, 38–46. [CrossRef]

22. Jiang, Y.; Zhou, Y.; Zhang, X. Cross-subject motor imagery EEG classification based on inter-domain Mixup fine-tuning strategy. *CAAI Trans. Intell. Syst.* **2024**, *19*, 909–919.