Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6, 2035-2054
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i6.8312

A more robust CNN-BiGRU-TPA model for wind turbine blade icing prediction

Tao Chen[1,2], Caixia Yang[1], Yao Xiao[1], Chaoying Yan[3], Chonlatee Photong[2*]
[1]College of Electrical Engineering, Hunan Mechanical and Electrical Polytechnic, Changsha 410073, China.
[2]Faculty of Engineering, Mahasarakham University, MahaSarakham 44150, Thailand; chonlatee.p@msu.ac.th (C.P.).
[3]College of Mechanical Engineering, Hunan Industry Polytechnic, Changsha 410208, China.

Abstract: Wind turbine blade icing poses a significant challenge to the reliability and efficiency of wind power generation, especially in cold and harsh climates. Accurately detecting icing conditions is essential for maintaining optimal turbine performance and preventing potential mechanical failures. However, conventional detection methods often face limitations when processing complex multivariate time-series data collected from Supervisory Control and Data Acquisition (SCADA) systems. In this study, we propose a novel hybrid deep learning model, CNN-BiGRU-TPA, which integrates Convolutional Neural Networks (CNNs) for spatial feature extraction, Bidirectional Gated Recurrent Units (BiGRUs) for temporal sequence modeling, and a Temporal Pattern Attention (TPA) mechanism to highlight critical temporal features. The proposed model demonstrates superior performance, achieving a recall of 0.8896, precision of 0.8904, F1 score of 0.8900, and Matthews Correlation Coefficient (MCC) of 0.7800. These results indicate its strong capability in accurately identifying icing events from complex SCADA datasets. This research provides a robust and scalable solution for real-time wind turbine monitoring and offers valuable insights for future applications in industrial big data analytics.

Keywords: *Bidirectional gated recurrent units, CNN-BiGRU-TPA model, Convolutional neural networks, Data analysis, SCADA data, Temporal attention mechanism, Wind turbine blade icing.*

1. Introduction

Wind power, as a clean and renewable energy source, has seen widespread adoption in recent years. It is progressively replacing fossil fuels and emerging as a crucial force in advancing carbon peak and carbon neutrality goals [1]. In China, with its vast territory and varied terrain, wind energy resources are abundant, particularly in high-altitude mountainous regions and areas near lakes in the south. However, these regions often experience cold temperatures and high humidity during winter, leading to frequent low temperatures and freezing conditions that can cause wind turbine blades to ice over. Blade icing not only reduces the efficiency of power generation but also increases the mechanical load on the blades, potentially resulting in serious safety hazards such as blade breakage or turbine failure. Therefore, the development of reliable ice detection methods is essential to ensure the safe and efficient operation of wind turbines.

At present, the diagnosis of wind turbine blade icing fault mainly relies on non-destructive testing technology, including blade icing detection system [2, 3] data mining method [4] acoustic emission detection [5] vibration mode recognition detection [6] infrared optical image processing [7] and unmanned aerial vehicle detection [8]. Among them, damage identification technologies such as ultrasonic [9] and fiber grating [10] are also applied in the inspection of wind turbine blades. Acoustic emission and vibration detection technologies often require sensors to be installed on the blades, which

History: Received: 24 April 2025; Revised: 26 May 2025; Accepted: 30 May 2025; Published: 23 June 2025
* Correspondence: chonlatee.p@msu.ac.th

is not only costly, but can also affect the dynamics of the blades. Drone detection faces the problems of downtime operation, short battery life and environmental factors. Hyper spectral imaging [11] and infrared cameras [12] can quickly and accurately identify the icing state of leaves through image processing, but the effectiveness of these methods may be limited in adverse weather conditions. The data mining method [13] performs well in accurately identifying the icing state of leaves, but it has high requirements for feature screening, and the accuracy of the overall algorithm may be affected if the feature screening is improper.

As the most widely used and mature data acquisition and monitoring system (SCADA), the SCADA system can provide rich information about the operating status of wind turbines without installing additional sensors, thereby reducing monitoring and maintenance costs, so it has been widely used in wind power equipment [14]. Machine learning and deep learning technologies are often used in the method of calculating the icing of wind turbine blades based on SCADA data, and the rapid development of these technologies has significantly promoted the research of ice detection of wind turbine blades.

Machine learning techniques are widely used in conventional wind turbine blade icing fault detection. Kreutz [15] proposed a convolutional neural network (CNN) model with dual inputs and a one-dimensional convolutional filter to effectively identify blade icing by utilizing historical turbine data and weather forecasts. Yang, et al. [16] developed a method using SCADA data, combining random forest and KNN algorithms to enhance feature selection, enhancing the accuracy of wind turbine blade icing diagnosis. This approach showed better performance compared to traditional BP neural networks, contributing to safer and more reliable wind turbine operation in cold regions. Xu, et al. [17] introduced a blade icing prediction method using a PSO-optimized SVM model. By enhancing data preprocessing and feature selection, the method improved prediction accuracy, offering an effective solution for icing fault detection. Meng, et al. [18] applied a combination of recursive feature elimination random forest (RFE-RF) and support vector machine (SVM) for icing monitoring, effectively diagnosing faults through feature selection and model fusion (Stacking). Tao, et al. [19] combined the physical characteristics of icing with the Stacked-XGBoost model to calculate wind turbine blade icing, significantly improving diagnostic accuracy and model generalization by accounting for both short-term and long-term icing effects.

Deep learning, due to its powerful ability to handle complex nonlinear problems, is widely used in the prediction of wind turbine blade icing, and there is a tendency toward the combination of multiple models. Xiong [20] based on the SCADA monitoring data of wind farms, a short-term icing state prediction model of wind turbine blades combining Bi-LSTM and SVM was proposed, and a high accuracy was achieved through PCA dimensionality reduction and multi-group data training. Dazhong [21] presented a deep fully connected neural network algorithm based on deep learning optimization, which shows higher prediction accuracy and computational efficiency in the prediction of icing state of wind turbine blades compared with the traditional KNN, SVM and unoptimized BP neural network methods. Tao [22] researched a method for predicting wind turbine blade icing employing a focal loss function combined with a CNN-Attention-GRU model. By refining feature extraction and handling data imbalance, their approach improved prediction accuracy by an average of 6.41% and increased the F1 score by 4.27%, significantly boosting the model's performance. In this study [23]. A leaf icing detection method was utilized, combining a spatiotemporal attention model with an adaptive weight (SAW) loss function. This approach significantly enhances classification accuracy and adaptability by extracting and weighting spatiotemporal features from imbalanced SCADA data while dynamically adjusting data category weights. The GTAN model [24] enhances the accuracy and robustness of wind turbine blade icing predictions by integrating feature extraction, temporal attention mechanisms, and a specialized loss function for handling imbalanced data, demonstrating strong potential for practical applications. Tian [25] constructed a multi-level convolutional recurrent neural network (MCRNN) with convolutional neural network (CNN) and long short-term memory network (LSTM) branches for leaf icing detection. Wang, et al. [26] offered a Wavelet LSTM method for detecting ice on wind

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

turbine blades. This approach incorporates wavelet-based multi-scale learning into the traditional LSTM framework, allowing for the simultaneous learning of both global and local temporal features from multivariable SCADA signals, thus improving fault detection performance.

Although existing methods for detecting ice formation in wind turbine blades, such as GSDE [27] ESS-ELM [28] and DCISS [29] model have made significant progress in improving the prediction accuracy and robustness, but there are still some problems such as insufficient generalization ability, strong dependence on high-quality labeled samples, and insufficient feature extraction. Especially when dealing with complex SCADA data, these models often ignore the temporal characteristics of the data and the dependencies between variables, leading to a reduction in detection accuracy in a dynamic environment.

Therefore, enhancing the accuracy of ice prediction of wind turbine blades, the paper constructs a hybrid prediction model combining Convolutional Neural Network (CNN) with Bidirectional Gated Recurrent Unit (BiGRU) network based on Temporal Pattern Attention Mechanism (TPA) [30, 31]. The model integrates the respective characteristics of CNN, BiGRU and TPA modules, and constructs a continuous feature map by inputting wind turbine operation data and environmental parameters through the prior knowledge of icing detection. Then, the CNN is used to extract the latent relationship between the time series data in the feature map, and the processed feature vector is used as the input of the BiGRU network. Finally, the dynamic timing modeling is carried out by combining the TPA mechanism to complete the short-term prediction of the icing of wind turbine blades. Compared to the BiGRU network model alone, the CNN-BiGRU-TPA hybrid model uses the CNN model to extract more important icing-related features. Additionally, the TPA module is incorporated into the BiGRU model to emphasize the key time series features after BiGRU analysis, thereby reducing the influence of insignificant features. This approach addresses the issue of the BiGRU model failing to distinguish the importance of time series features, resulting in improved prediction accuracy and robustness.

2. Related Work

Icing of wind turbine blades is often regarded as a multivariate time series binary classification problem. In the day-to-day operation of a wind farm, it is critical to quickly determine whether the blades are icy so that appropriate measures can be taken in a timely manner, such as stopping the aircraft or starting the de-icing equipment. By simplifying the problem into two categories, "icing" and "non-icing", the binary classification model not only improves detection efficiency and accuracy, but also simplifies the data processing process, which helps to achieve faster real-time responses. In addition, the binary classification method requires fewer computing resources and is particularly suitable for the diagnosis of ice formation of wind turbine blades in large-scale wind farms. The data-driven approach to modeling wind turbine blade icing primarily involves feature engineering, distance analysis, and deep learning techniques. Feature engineering is concerned with extracting key features from the time, frequency, and time-frequency domains to facilitate the classification of time series data.The distance analysis method distinguishes between normal and abnormal samples by calculating the distance between samples. However, the limitation of these methods is that they require expertise to select and extract useful features and suitable distance metrics, which often require a significant investment of resources. In contrast, deep learning, as a data-driven approach, can overcome these limitations and provide more effective solutions.

2.1. CNN Neural Network

The Convolutional Neural Network (CNN, as shown in Figure 1) is a deep learning model commonly applied to computer vision tasks, especially for image recognition and classification involve CNNs that extract features from image data through convolution and pooling operations, followed by classification or regression in fully connected layers. The main components of a CNN are the convolutional and pooling layers. CNN [32] has shown excellent performance in solving many

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

problems related to multi-dimensional nonlinear data and has been extensively used in the prediction of wind turbine blade icing.

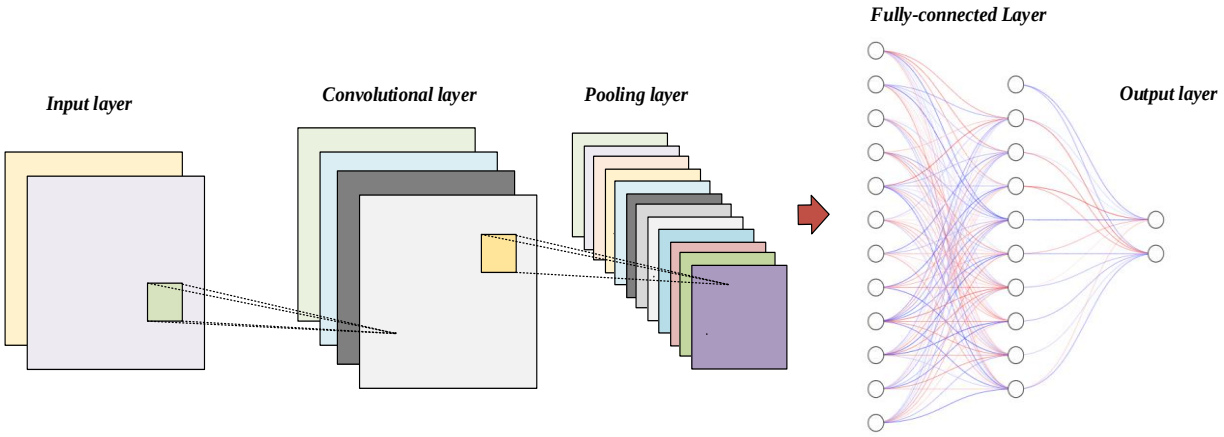


Figure 1.
Framework of CNN.

The convolutional layer operation usually uses a filter-like convolutional kernel mechanism to extract the features of the icing parameter matrix of wind turbine blades in the SCADA system, extracts data blocks from the input feature vectors, transforms them, and then reorders the output feature vectors spatially. This operation reduces the quantity of parameters in the overall network, prevents overfitting from occurring, and reduces the amount of memory occupied by the entire network, thus reducing the amount of computation. In practice, convolution operations are usually used instead of convolution operations because they need to flip the convolution kernel during backpropagation, as shown in equation (1):

$$y^{l(i,j)} = K_i^l * X^{l(r^j)} = \sum_{j'=0}^{W-1} K_i^{l(j')} X^{l(j+j')} \quad (1)$$

Where $K_i^{l(j')}$is the j' weight of the i^{th} convolutional kernel of l-layer;$X^{l(j+j')}$ represents the j^{th} convoluted local region of the i^{th} convolution kernel of l-layer; W represents the width of the convolution kernel.

The pooling layer is mainly reducing parameters through downsampling, and the pooling process is described as follows:

$$A_k^l(i,j) = \left[\sum_{x=1}^{f}\sum_{y=1}^{f} A_k^l\,(s_0 i + x, s_0 j + y)^p\right]^{\frac{1}{p}} \quad (2)$$

Where s_0 represents the step size, f represents the size of the convolution kernel, p fills the number of layers, and when $p \to \infty$, the maximum value is taken in the pooling area, that is, the maximum pooling. The pooling layer optimizes the dimensionality by selecting the most representative features to optimize the network. However, each spatial position in the output feature vector matches to the same position in the input feature vector, that is, the original sequential relationship is still retained after the convolutional pooling operation, and the temporal characteristics of the leaf icing sequence are not destroyed.

2.2. GRU

A Recurrent Neural Network (RNN) is a model designed to process time series data by retaining previous information through the interconnected structure of nodes at each layer. This stored information is then used to influence the output of subsequent layers, enabling the model to capture temporal dependencies in the data.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

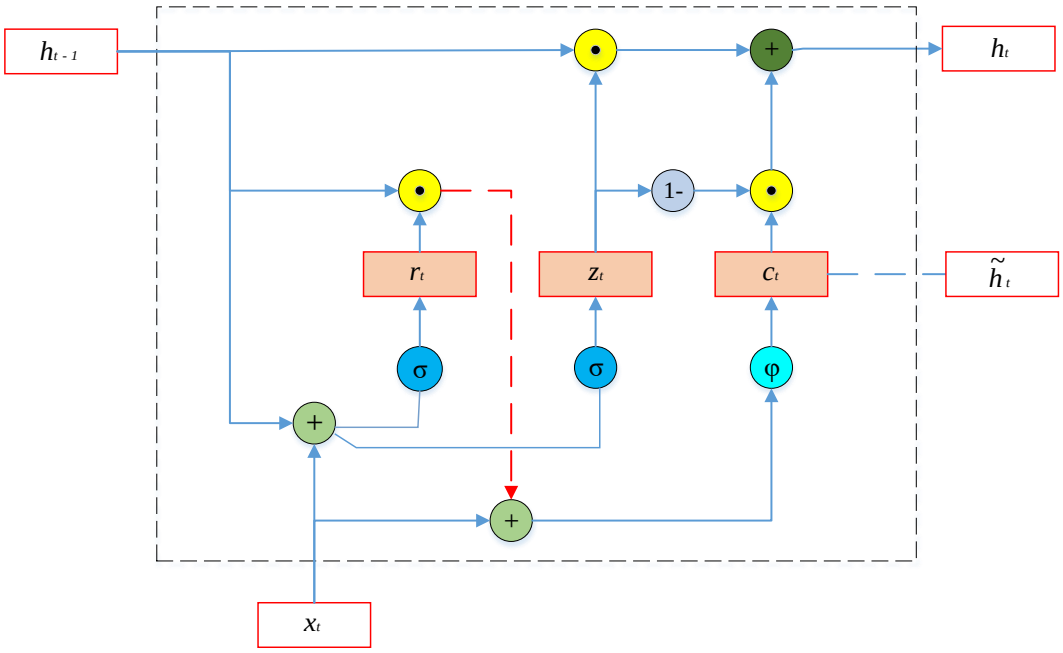


Figure 2.
Framework of GRU.

The original Recurrent Neural Network (RNN) struggles with learning long-term dependencies in sequential data and is susceptible to issues like vanishing or exploding gradients, making gradient-based optimization methods challenging. To address these problems, two extended models were proposed: Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). GRU, a simplified version of LSTM, features a more efficient structure. As depicted in Figure 2, GRU [33] is composed of three parts: the reset gate (r_t), the update gate (z_t), and the candidate layer (c_t). Unlike LSTM, GRU eliminates the need for an additional memory cell and uses the update gate (z_t) to regulate which information is retained or discarded, thus reducing computational overhead. The GRU's operations can be computed as follows:

$$z_t = \sigma\left(V_{xz}x_t + U_{hz}h_{t-1} + b_z\right) \quad (3)$$

$$r_t = \sigma\left(V_{xr}x_t + U_{hr}h_{t-1} + b_r\right) \quad (4)$$

$$c_t = \varphi\left(V_{xc}c_{xt} + U_{hc}\left(r_t \odot h_{t-1}\right) + b_c\right) \quad (5)$$

Where V_{xz}, V_{xr}, V_{xc} denote the weights that link the input layer to the GRU units. U_{hz}, U_{hr}, U_{hc} represent the weights associated with the self-connections between the current time step t and the previous time step $t-1$.b_z , b_r denote the biases for the two gates in the GRU unit, while b_c represents the bias for the candidate layer.

2.3. Bidirectional GRU

While GRU is effective in modeling time-series data, it is restricted to learning information solely in the forward direction. In contrast, the Bidirectional GRU (BiGRU) improves upon this by incorporating two layers-one that processes information in the forward direction and another in the backward direction. As illustrated in Figure 3, the forward and backward layers are connected to the output layer, allowing the model to capture both past and future context, making it more effective in processing sequential data with dependencies across different time steps.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

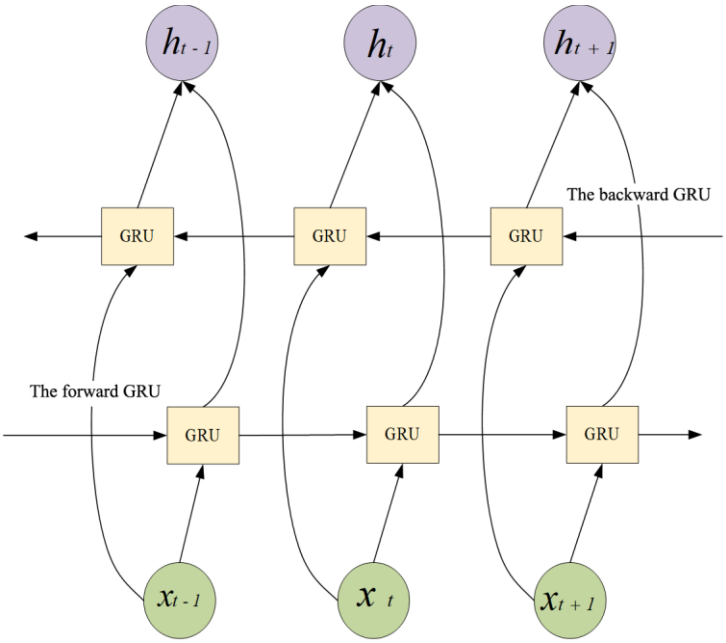


Figure 3.
Framework of BiGRU.

The forward layer ($\vec{h}_t$) conducts training on the time-series data in the forward direction, while the backward layer ($\overleftarrow{h}_t$) trains the time-series data in the reverse direction.This bidirectional approach allows the model to utilize information from both directions, enhancing its understanding of the sequence. The forward and backward training processes can be formulated as follows:

$$\vec{h}_t = f(\vec{W} \cdot x_t + \vec{U} \cdot \vec{h}_{t-1} + \vec{b}) \quad (6)$$

$$\overleftarrow{h}_t = f(\overleftarrow{W} \cdot x_t + \overleftarrow{U} \cdot \overleftarrow{h}_{t-1} + \overleftarrow{b}) \quad (7)$$

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (8)$$

The weights connecting the input layer to the forward and backward layers are denoted by $\vec{W}$and $\overleftarrow{W}$, respectively. The self-connection weights for these layers are represented by $\vec{U}$and $\overleftarrow{U}$. The biases for the forward and backward layers are $\vec{b}$ and $\overleftarrow{b}$, respectively. Additionally, $\oplus$ represents the concatenation of the outputs from both layers.

2.4. Temporal Attention Mechanism

The attention mechanism Yin and Jesse [34] and SHI, et al. [35] mimics human visual attention by assigning different weights to various features, emphasizing those most relevant to the target task. Typical attention mechanisms focus on extracting important information associated with the current time step. The context vector v_t is derived by weighting the column vectors of the BiGRU hidden state $H=(h_1, h_2, ..., h_{t-1})$. However, in the case of wind turbine blade icing detection, each time step involves multiple variables that exhibit complex nonlinear relationships. Additionally, each variable exhibits unique characteristics and periodicity, making it challenging to pinpoint a specific time step as the primary focus. To address this, the Temporal Pattern Attention Mechanism (TPA) applies a one-dimensional convolutional neural network (1D-CNN) filter to extract row vector features from the BiGRU hidden state *H*. This improves the model's capacity to capture the relationships among multiple variables across various time steps. The framework of TPA is shown in Figure 4.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

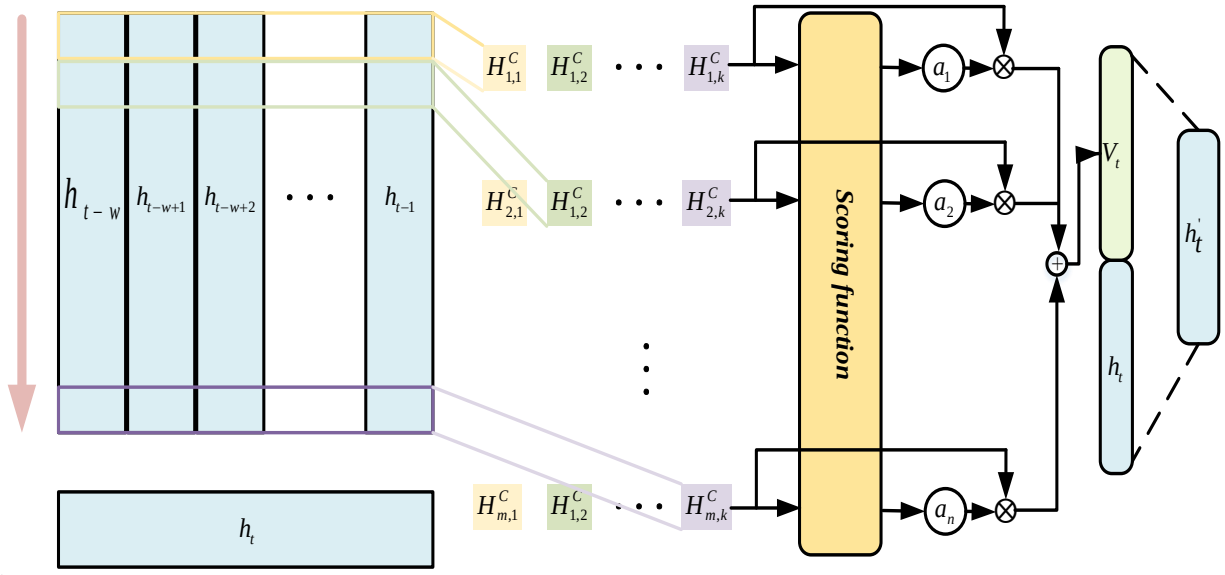


Figure 4.
Framework of TPA.

For the initial time series H, BiGRU computes a hidden state matrix $H_{i,j}$, and TPA applies k filters to extract features through row-wise convolution operations on each row H_i of H, generating a time pattern matrix $H_{i,j}^C$. Here, H_{ij} represents the output value corresponding to the i-th row vector and the j-th convolutional kernel, with C representing the convolution process.

$$H_{i,j}^C = \sum_{l=1}^{w} H_{i,(t-w-1+l)} \times C_{j,T-w+l} \quad (9)$$

The predicted nodes are h_t and $H_{i,j}^C$. Each matrix row is processed to calculate a weight, which is then normalized to derive the attention weight a_i, reflecting the influence of each time series on h_t.

$$f(H_i^C, h_t) = (H_i^C)^T W_a h_t \quad (10)$$

$$a_i = \text{sigmoid}(f(H_i^C, h_t)) \quad (11)$$

The weighted sum of the weights of each time series is used to obtain the weight vector v_t, which takes into account the combined effect of all rows on h_t, i.e., time attention.

$$v_t = \sum_{i=1}^{n} a_i H_i^C \quad (12)$$

For v_t and h_t, the linear mapping and the sum are performed to obtain the time information state h_t'.

$$h_t' = W_t h_t + W_v v_t \quad (13)$$

Where $W_t \backslash W_v$ represents the weight matrix of different variables.

A key advantage of TPA over traditional attention mechanisms is its ability to assign weights not only to time steps, but also to individual characteristic dimensions related to icing of wind turbine blades. Standard attention mechanisms often distribute weights between different time steps, which makes it difficult to pinpoint the specific effect of each variable on leaf icing within a single time step. In contrast, TPA digs deep into the relationship between variables in historical data and current icing conditions, using weighted attention vectors to better capture time series features to generate more accurate predictions. In addition, while RNN models such as BiGRU often encounter problems such as vanishing gradients when dealing with long-term wind turbine SCADA data series, TPA overcomes this challenge by analyzing the correlation between historical icing sequences and future results through its resource allocation system. This strategy emphasizes key contributors, reduces the risk of losing important information, and effectively captures long-term dependencies, making TPA a powerful

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

tool for predicting icing on wind turbine blades, and Table 5 can demonstrate the superiority of its model.

2.5. CNN-BIGRU-TPA Model

Figure 5 shows the architecture of the CNN-BiGRU-TPA model developed in this study, consisting of five main components: the input layer, CNN layer, BiGRU layer, TPA layer, and output layer. The roles of these layers are as follows:

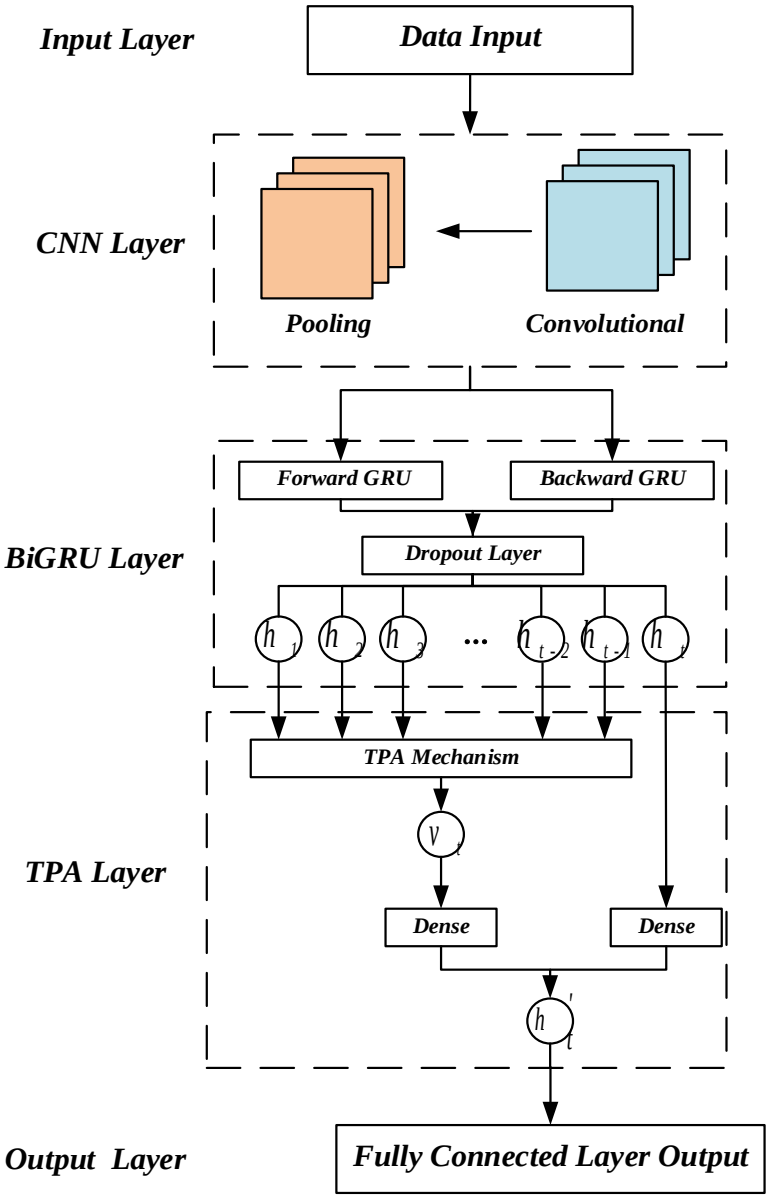


Figure 5.
Framework of CNN-BIGRU-TPA.

(1) Input Layer: The model's input consists of historical SCADA icing data and other influencing factors.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

(2)CNN Layer: This layer captures the relationships between input factors and current blade icing. The convolutional layers identify feature patterns, while the pooling layers reduce dimensionality by selecting key features. The processed features are then sent to the BiGRU layer via fully connected layers.

(3) BiGRU Layer [36]: The BiGRU layer captures temporal dependencies within the icing sequence, building a dynamic time-series model. It processes bidirectional information and applies dropout regularization to prevent overfitting. The result is a hidden state matrix $H_{ij} = [h_1, h_2, ..., h_{t-1}, h_t]$, which is passed to the TPA layer.

(4) TPA Layer: The Temporal Pattern Attention Mechanism(TPA) assigns adaptive weights to the hidden states from BiGRU, selecting and emphasizing relevant features. The attention mechanism produces a comprehensive temporal feature vector h_t' by highlighting important features.

(5) Output Layer: The fully connected layer produces the predicted output, indicating whether blade icing will occur in the next time step.

3. Experiments and Discussion

3.1. Data Description

In this study, icing data for wind turbine blades was collected from a wind farm located in Hunan Province, utilizing turbines manufactured by Harbin electric corporation. The SCADA data was recorded from December 1st to December 30th, with a resolution of 7 seconds. Engineers identified 26 key variables related to blade icing from a SCADA system with data from hundreds of sensors and pinpointed the times when icing occurred. These variables are listed in Table 1.

Table 1.
Partial Monitoring Parameter Information of Wind Turbine Units.

Serial Number	Variable Name	Serial Number	Variable Name
1	Wind_speed	14	Pitch1_moto_tmp
2	Generator_speed	15	Pitch2_moto_tmp
3	Power(kw)	16	Pitch3_moto_tmp
4	Wind_direction(°)	17	Acc_xx
5	Wind_direction_mean	18	Acc_yy
6	yaw_position	19	environment_tmp
7	yaw_speed	20	int_tmp
8	Pitch1_speed	21	Pitch1_ng5_tmp
9	Pitch2_speed	22	Pitch2_ng5_tmp
10	Pitch3_speed	23	Pitch3_ng5_tmp
11	Pitch1_angle	24	Pitch1_ng5_DC
12	Pitch2_angle	25	Pitch2_ng5_DC
13	Pitch3_angle	26	Pitch3_ng5_DC

3.2. Data Preprocessing

The data used in this study was obtained from the SCADA system, which presented challenges such as outliers, redundant entries, and missing values. To prevent these issues from impacting model training and testing, a comprehensive data preprocessing procedure was carried out, including the following steps:

(1) Read and preliminarily filter data: Use the os.listdir() function to iterate through all the files in the data directory, perform preliminarily filter based on specific keywords in the file name, and then read the CSV file that matches the criteria.

(2) Data labeling and imputation: Experienced wind turbine engineers labeled all raw data, categorizing it as either normal or iced. To ensure data continuity and completeness, missing values were handled using the Fillna method, preventing gaps in the time series from affecting subsequent analysis and model training.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

(3) Data segmentation and normalization: Since the SCADA data is collected as time series from sensors at regular intervals, normalization and standardization were applied to align the data with the input requirements of the CNN-BiGRU-TPA wind turbine blade icing prediction model. First, Min-Max normalization was performed to scale the feature values within the range of 0 to 1, mitigating the impact of varying units and magnitudes on the model. The normalized time series was then segmented into fixed-length fragments (500 data points) suitable for model input.

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (14)$$

Here, X' represents the preprocessed feature data, X is the raw data, and X_{max} and X_{min} represent the maximum and minimum values of the feature, respectively.

(4) Handling iced data: In addition to processing the normal data, similar segmentation and normalization procedures are applied to the iced data. Failure data is integrated into the overall dataset by retrieving information from other data sources and applying the same preprocessing techniques.

3.3. Evaluation Metrics

This study evaluates the proposed model using imbalanced data, where the number of abnormal (icing) samples is much smaller than the number of normal (non-icing) samples. This imbalance can inflate the precision score, making it unsuitable for accurate model assessment. Given the focus on wind turbine blade icing, with icing and non-icing conditions treated as positive and negative samples, respectively, the paper uses precision, recall, F1-score, and the Matthews correlation coefficient (MCC) as performance metrics. The definitions of these metrics are as follows:

$$Pr\,e\,cision = \frac{TP}{TP+FP} \quad (15)$$

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

$$F1 = \frac{2 \times Pr\,ecision \times Re\,call}{Pr\,ecision + Re\,call} \quad (17)$$

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \quad (18)$$

$$MCC = \frac{TP \times TN - FP \times FN}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (19)$$

Where TP, FP, FN, and TN represent true positives, false positives, false negatives, and true negatives, respectively.

3.4. Experimental Environment

This experiment was performed in a computing environment with the following Table 2:

Table 2.
Experimental Environment.

Hardware configuration	Software Configuration
(1)CPU: Intel64 Family 6 Model 186 Stepping 2, GenuineIntel (2)Memory: 31.65 GB RAM (3)Storage: 302.60 GB	(1)Operating System: Windows 10 10.0.22631 SP0 (2)Programming language: Python 3.8 (3)Deep learning framework: PyTorch 1.8.1

4. Case Study

4.1. Data Description

4.1.1. LSTM and BiLSTM

The LSTM model was used to capture unidirectional temporal features for diagnosing wind turbine blade icing, while the BiLSTM model extracted both forward and backward temporal information, improving the overall understanding of blade conditions. Both models analyzed and classified icing states using a cross-entropy loss function (which implicitly included a softmax classifier).

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

4.1.2. GRU and BiGRU

The GRU model was used to capture unidirectional temporal features in wind turbine blade icing diagnosis, as shown in Section 2.2. BiGRU, on the other hand, uses a bidirectional structure to extract both forward and backward temporal information simultaneously, as shown in Section 2.3, enhancing the overall understanding of blade conditions. Both models incorporated a temporal attention mechanism to focus on key time steps and analyze icing states using a cross-entropy loss function (which implicitly included a softmax classifier).

4.1.3. CNN and CNN-BiGRU-TPA

A CNN layer and the CNN-BiGRU-TPA model were used with the parameters and structure described in Section 2.1 and 2.5.

4.2. Baseline Comparison

To determine the optimal model parameters, a controlled variable method was used to gradually adjust and refine the hyperparameters. Initially, the model was set with 100 training epochs, a batch size of 100, a learning rate of 0.001, and 100 hidden units for BiGRU. The convolutional kernel size and the number of filters in the CNN were then adjusted. The results showed that when the kernel size was 3 and the number of filters was 32, the model's precision and accuracy were relatively low. The performance of the CNN with 2 and 3 layers was also tested, but adding more layers did not enhance prediction accuracy, so a single-layer CNN was selected.

After determining the parameters for the CNN layer, the hidden units and number of layers for the BiGRU were optimized. It was found that setting the hidden units to 64 and using 1 BiGRU layer minimized the prediction error. Next, with the CNN and BiGRU hyperparameters fixed, a dropout layer was added after BiGRU, and a dropout rate of 0.2 was found to give the best performance.

The hyperparameters of each model were mainly determined by factors such as: initial learning rate (ILR), number of hidden units (HUs), learning rate decay mechanism (using CosineAnnealingLR), and the number of filters in the CNN layers (FILs). The learning rate for each model was fine-tuned within the range of [0.001, 0.005]. All models used L2 regularization, controlled by the weight decay, with values between 0.0005 and 0.001. To reduce the risk of overfitting, dropout settings varied across models, with the BiLSTM model employing higher dropout rates (0.2/0.3). The number of hidden units was set between , and for CNN-related models, multiple kernel sizes (5, 7, 9) were used to capture multi-scale features.

All models used a consistent batch size of 512 and cross-entropy as the loss function. The table 3 below shows the final hyperparameter values that achieved the best performance for each model, representing the optimal configuration for both performance and optimization.

Table 3.
Partial Monitoring Parameter Information of Wind Turbine Units.

Model	Learning Rate	Weight Decay	Dropout	Hidden Size	Filter Sizes	Batch Size
CNN-TPA	0.0003	0.0005	0.1	N/A	[5, 7, 9]	512
LSTM-TPA	0.0003	0.0005	0.1	64	N/A	512
GRU-TPA	0.0003	0.0005	0.1	64	N/A	512
BiGRU-TPA	0.0003	0.0005	0.1	64/128	N/A	512
BiLSTM-TPA	0.0001	0.001	0.2	64/128	N/A	512
CNN-BiGRU-TPA	0.0003	0.0005	0.1	64/128	[5, 7, 9]	512

4.3. Baseline Comparison

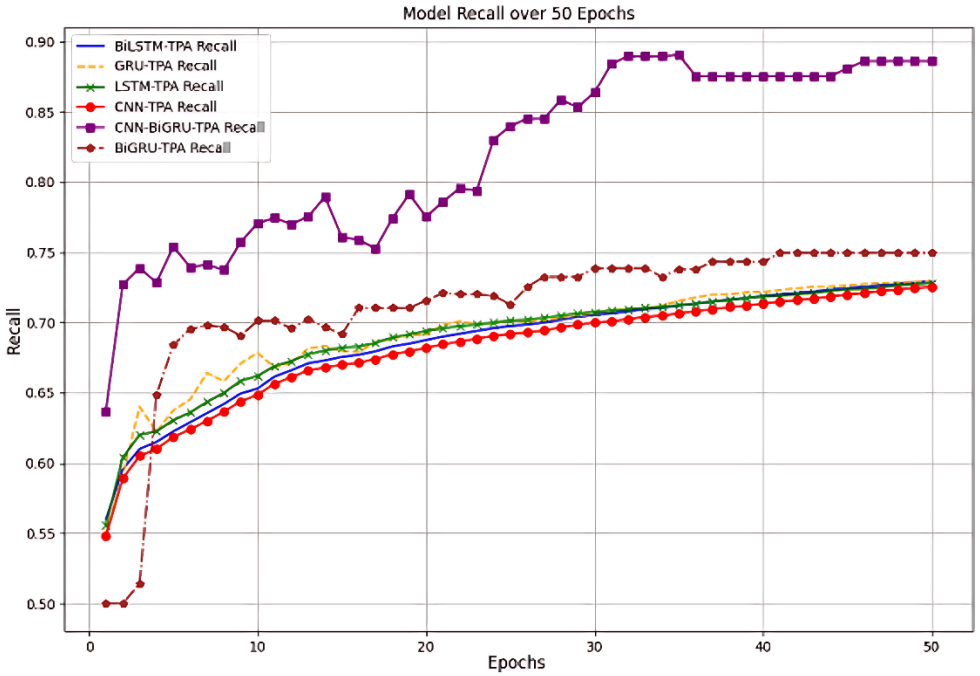
Figure 6 shows the performance of different models during training across 50 epochs, highlighting key metrics. From the data, we studied the behavior of each model, focusing on the jumping phenomenon, flattening trends, and why some models show less significant loss reduction.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

4.3.1. Jumping Phenomenon

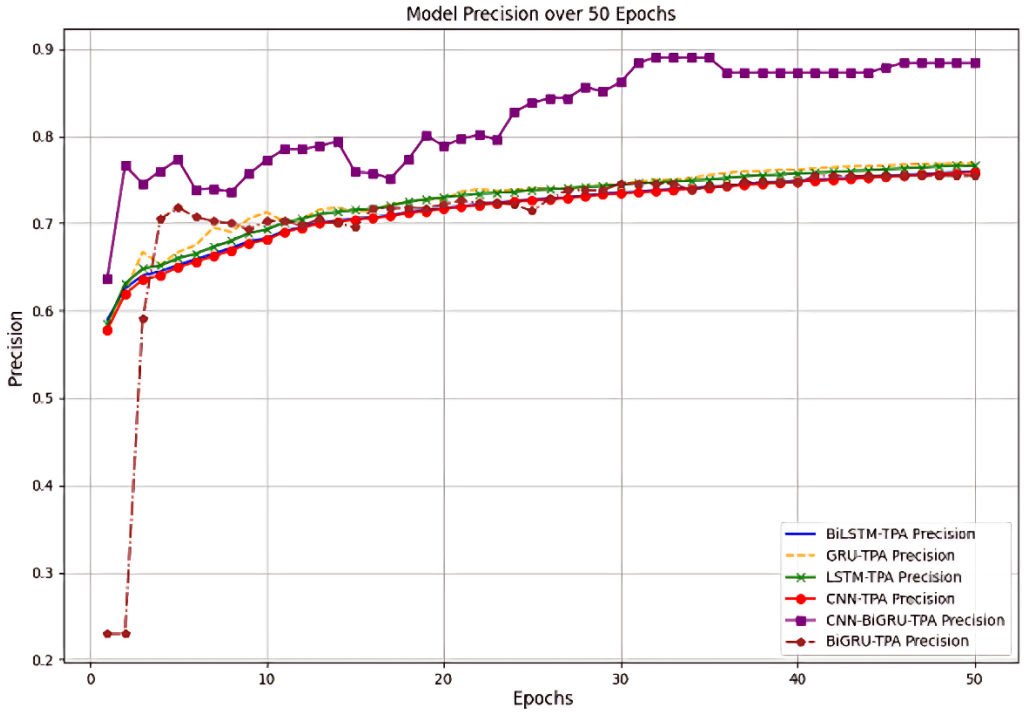
The curves of the BiGRU-TPA model on multiple indicators, such as Precision (Figure 6(b)) and MCC (Figure 6(d)), showed obvious jumps. This phenomenon might be due to the model's sensitivity to the data, especially in the early training stages, when the model weights were updated drastically, resulting in large fluctuations in the metrics. Since the BiGRU model considered both forward and backward information, if the attention mechanism of the model was unstable, it would cause large changes in the weights of local time steps, leading to fluctuations in prediction accuracy.

Additionally, the jump in the CNN-BiGRU-TPA model was also reflected in the early stage, primarily due to the complexity of the model structure. The model required to balance between extracting local features in the convolutional layer and global temporal features in BiGRU. Early in training, the model might be trying to optimize the weights of these two components, causing the metrics to fluctuate greatly.

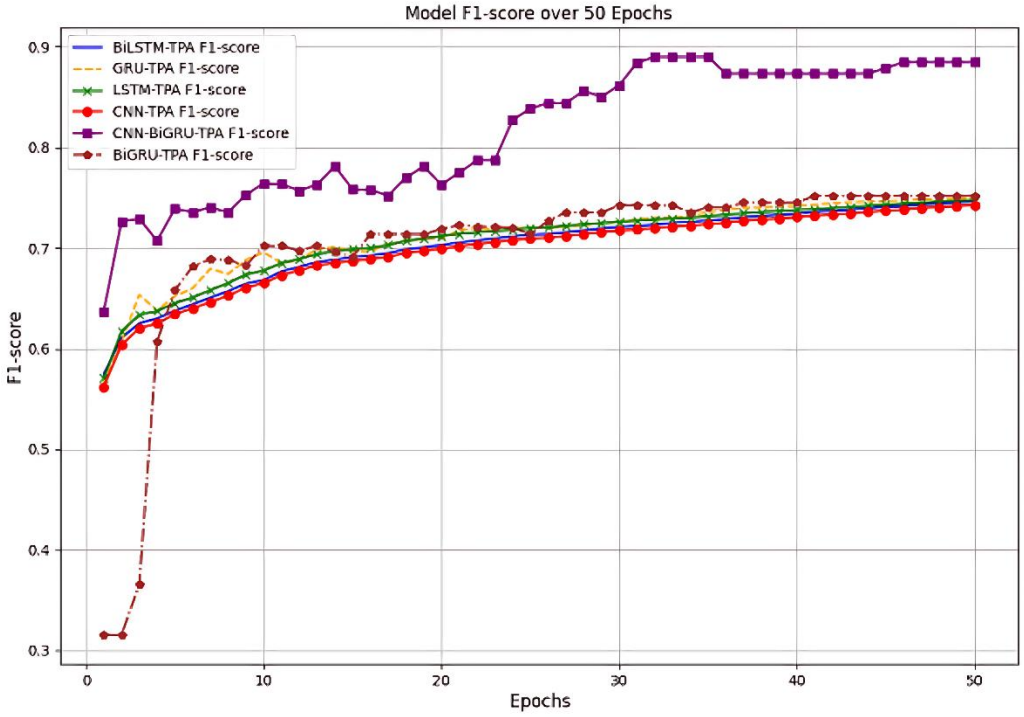


(a) Recall

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

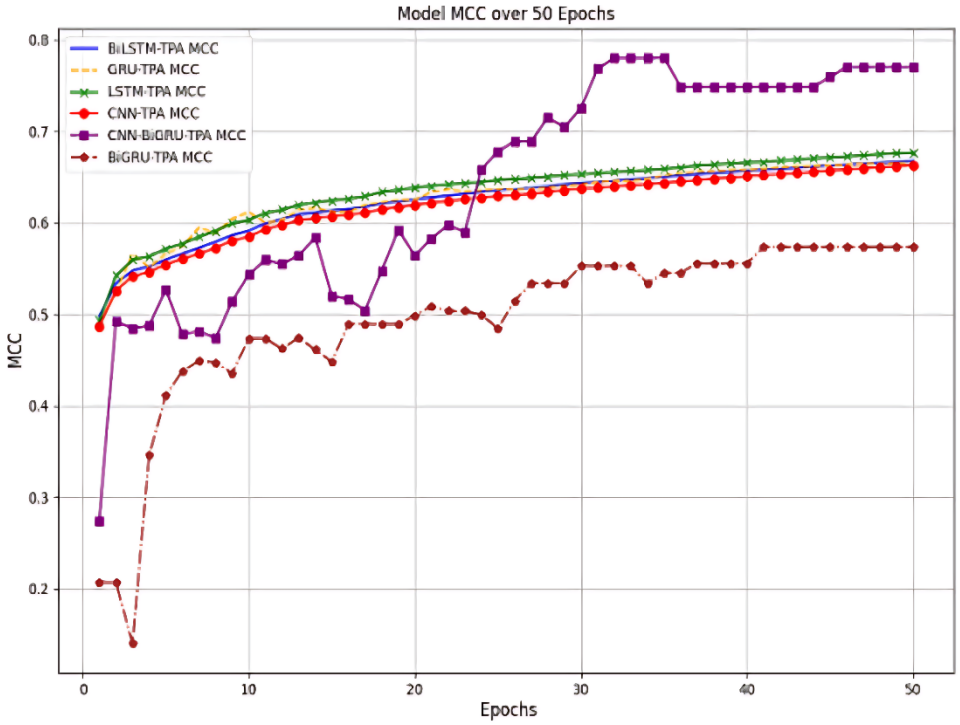


(b) Precision

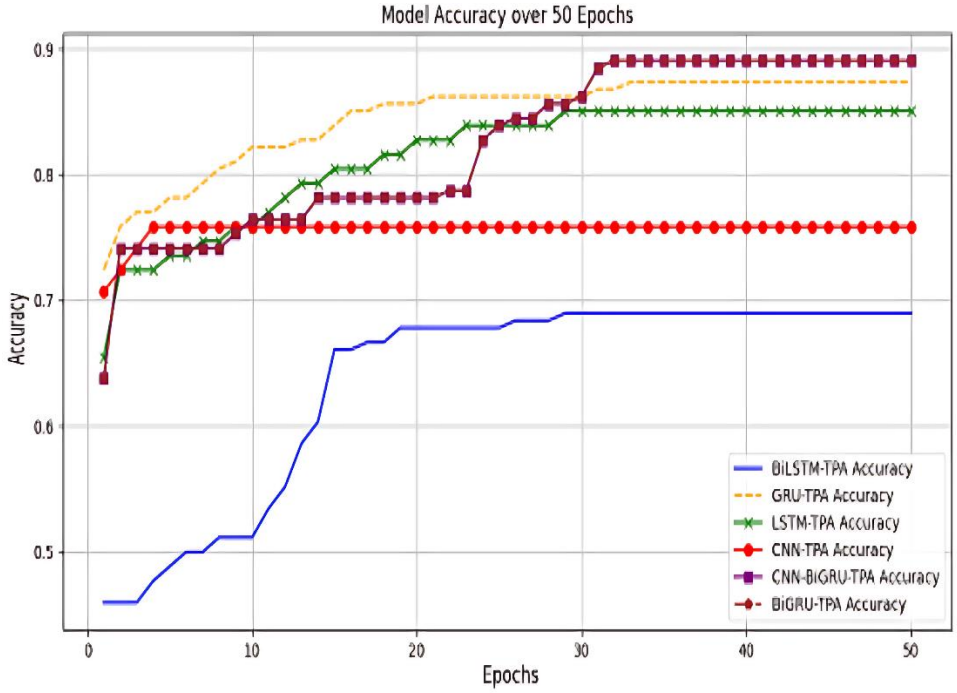


(c) F1-score

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

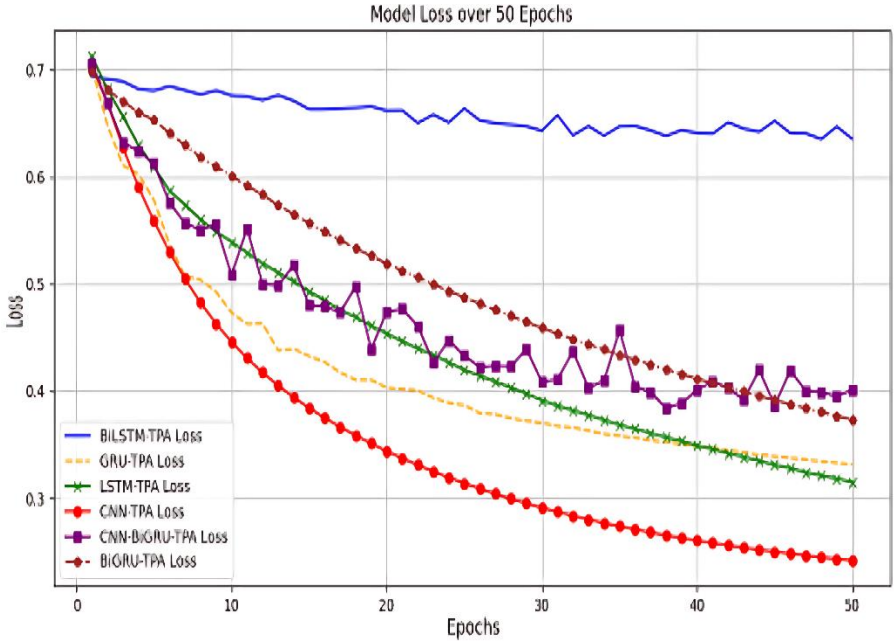


(d) MCC



(e) Accuracy

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312



(f) Loss

Figure 6.
Performance Comparison of the models over 50 epochs.

4.3.2. The Flattening Phenomenon

The index curves of the LSTM-TPA and GRU-TPA models were relatively stable, as seen in Figure 6(a) and 6(b). This might be because these models had simpler structures and focus on one-way temporal features. The optimization direction of these models remained relatively stable in each training cycle, resulting in a slow, steady improvement of the metrics with small fluctuations.

The curve of the CNN-TPA model also showed a relatively flat trend, especially the decrease in the Loss (Figure 6f). After about 30 epochs, the loss value decreased by approximately 20%. This indicated that the convolutional network's feature extraction was relatively stable, and the model gradually converged during the training process.

4.3.3. The loss of the CNN-BiGRU-TPA Model

Although CNN-BiGRU-TPA excels in terms of Accuracy (Figure 6(e)), Recall (Figure 6(a)), and F1 score (Figure 6(c)), its Loss curve (Figure 6(f)) would not decrease significantly in the later stages of training (between 30 and 50 epochs). Several reasons might account for this:

Complexity of the model: CNN-BiGRU-TPA combined convolutional networks and bidirectional GRUs, along with a time-attention mechanism, resulting in a large number of parameters during training. While this complex structure could be able to extract richer features, it also caused the model to fall into a local optimum, preventing further significant reductions in the loss value.

The flat and jumping changes in different model curves were closely related to the complexity of the model, optimization difficulty, and the structural characteristics. Although CNN-BiGRU-TPA exhibited superior classification performance, the model's complexity caused the loss value to plateau, reflecting the model's tendency to converge to a local optimum during optimization.

4.4. Baseline Comparison

Table 4 shows the performance of six models in detecting wind turbine blade icing, evaluated on five key metrics: Recall, Precision, Accuracy (ACC), F1 Score (F1), and Matthews Correlation

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

Coefficient (MCC). A comparison of these metrics indicates that the CNN-BiGRU-TPA model outperforms the others.

Table 4.
Classification results of six models.

Models	Recall	Precision	ACC	F1	MCC
CNN-TPA	0.7967	0.8051	0.7874	0.7868	0.6017
LSTM-TPA	0.7807	0.7929	0.7701	0.7690	0.5735
GRU-TPA	0.7210	0.7263	0.7126	0.7120	0.4473
BiLSTM-TPA	0.7044	0.7284	0.6897	0.6846	0.4322
BiGRU-TPA	0.7495	0.7544	0.7414	0.7410	0.5039
CNN-BiGRU-TPA	0.8896[1]	0.8904[1]	0.8908[1]	0.8900[1]	0.7800[1]

From the perspective of Accuracy, F1-Score, Precision, Recall, and MCC, the CNN-BiGRU-TPA model showed obvious comprehensive advantages. First, on the Accuracy, CNN-BiGRU-TPA reached 0.89081, which was about 20%-25% higher than the 0.7414 and 0.7126 of BiGRU-TPA and GRU-TPA, showing a significant improvement in classification accuracy. On F1-Score, CNN-BiGRU-TPA reached 0.89001, an improvement of about 10% over other models, especially compared to BiLSTM-TPA's 0.6846. At the same time, for Precision and Recall, the CNN-BiGRU-TPA achieved 0.89041 and 0.88961, respectively, much higher than other models, demonstrating excellent balance when detecting positive and negative samples. In addition, in the MCC indicator, CNN-BiGRU-TPA reached 0.78001, which was 36% higher than that of LSTM-TPA of 0.5735, indicating that it was more robust in dealing with the classification task of unbalanced data.

In conclusion, the CNN-BiGRU-TPA model significantly improved the classification performance by combining the feature extraction capabilities of CNN and the temporal information capture advantages of BiGRU. In contrast, the indicators of LSTM-TPA and GRU-TPA were slightly lower than those of the CNN-TPA models, especially on Precision and Recall, both of which did not reach the performance level of CNN-BiGRU-TPA, suggesting that they had some errors and limitations in detecting positive and negative samples of wind turbine blade icing. On the whole, the improvement of CNN-BiGRU-TPA in a number of indicators proved its absolute advantage in processing complex time series data and unbalanced data, especially for task scenarios that require high-precision classification.

4.5. Comparison to Different Optimization Approaches

To further demonstrate the superiority of the CNN-BiGRU-TPA model, this paper compares models without attention mechanisms (such as MLP and ResNet) with models incorporating attention modules (such as SE, GA-BiGRU-TPA, and GWO-BiGRU-TPA). The evaluation metrics are shown in Table 5.

Table 5.
Evaluation metrics of different models.

Models	Recall	Precision	F1	MCC	References
MLP	0.607	0.718	0.714	0.420	[30]
ResNet	**0.885**[3]	0.778	0.834	0.674	[30]
SE	0.7662	**0.8215**[2]	0.7924	0.7778	[29]
GTAN	0.794	0.808	0.788	0.754	[23]
GA-BiGRU-TPA	0.8735	**0.8179**[3]	**0.8443**[3]	**0.8326**[2]	[29]
GWO-BiGRU-TPA	**0.9158**[1]	0.8141	**0.8618**[2]	**0.8523**[1]	[29]
CNN-BiGRU-TPA	**0.8896**[2]	**0.8904**[1]	**0.8900**[1]	**0.7800**[3]	Ours

The MLP model Yuan, et al. [37] consists of three fully connected (FC) layers, each with 500 hidden nodes, and dropout layers between them, serving as a basic baseline model. On the other hand, ResNet [38] is a widely used CNN variant in time series classification (TSC) tasks, with a complex and

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

deep architecture capable of extracting rich features. However, both of these models do not incorporate temporal attention (TA) mechanisms, limiting their baseline performance.

To address this, the SE model [39] incorporated the Squeeze-and-Excitation (SE) module, which captures channel interdependencies and adaptively adjusts the feature responses. Additionally, the GTAN model introduces a feature extraction module designed to enhance the differentiation between various sensor data categories while integrating a temporal attention (TA) mechanism to improve sensitivity to temporal features. The inclusion of these attention mechanisms enabled the models to handle complex time series data more effectively.

The GA-BiGRU-TPA model optimized the hyperparameters of the Bidirectional Gated Recurrent Unit (BiGRU) through genetic algorithm (GA) [40] and combined it with the Time Attention Mechanism (TPA) to enhance the focus on key time steps and improved the prediction ability of complex time series data. Similarly, the GWO-BiGRU-TPA model optimizes the BiGRU hyperparametrically through the Gray Wolf Optimizer (GWO) [41] and combines it with TPA, which further improved the model's ability to capture key temporal features and exhibits higher prediction accuracy.

The CNN-BiGRU-TPA model outperformed other models in key performance metrics by combining CNN's spatial feature extraction, BiGRU's ability to capture time dependence, and TPA's enhanced weight allocation across time steps. For instance, CNN-BiGRU-TPA achieved a Precision of 0.89041 and a Recall of 0.88962, significantly surpassing the SE model (Precision: 0.8215, Recall: 0.7662). This highlights the effectiveness of integrating spatial and temporal attention mechanisms. It also outperformed optimization-based models like GA-BiGRU-TPA (Precision: 0.81792, Recall: 0.8735) and GWO-BiGRU-TPA (Precision: 0.8141, Recall: 0.9158), demonstrating more consistent and superior performance across multiple metrics.

By combining CNN's local spatial feature extraction, BiGRU's bidirectional time-series capture, and an optimized temporal attention mechanism, CNN-BiGRU-TPA outperforms traditional models and other attention-based approaches in wind turbine blade ice detection, making it ideal for complex time series prediction tasks.

5. Case Study

This study proposes significant advancements in the detection accuracy of wind turbine blade icing through the development and refinement of the CNN-BiGRU-TPA model. By incorporating CNN for spatial feature extraction and BiGRU to capture temporal dynamics in multivariate time-series data, the model utilizes the Temporal Pattern Attention Mechanism(TPA) to highlight crucial time series information. The proposed model showed to outperform both traditional and enhanced models across various key performance indicators. Specifically, the proposed CNN-BiGRU-TPA model achieved a recall of 0.8896, precision of 0.8904, an F1 score of 0.8900, and an MCC of 0.7800, indicating its high predictive accuracy and robustness.

The model also incorporated min-max normalization and time window segmentation during data preprocessing, ensuring that the input data was consistent and standardized. Results from tests conducted in a high-performance computing environment indicate that the model not only improved accuracy but also speeded up convergence, reaching a stable accuracy of 0.89 after just 30 training cycles.

In summary, the proposed CNN-BiGRU-TPA model showed exceptional performance in identifying icing on wind turbine blades, offering a novel approach to tackling complex industrial big data problems. The model's success suggested considerable potential for broader applications. Future research should aim to refine the model structure and explore new approaches for data augmentation and model integration to improve prediction accuracy and expand its applicability.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Acknowledgement:

Thanks to Hunan Mechanical & Electrical Polytechnic and the Faculty of Engineering, Mahasarakham University, for their support, especially to Dr. Chonlatee Photong for his endless support. Special thanks are also extended to the Institute of Energy Power and Fault Diagnosis Technology, Hunan Mechanical & Electrical Polytechnic, for providing valuable support and resources for this research (Project No. YJS24-01).

References

[1] Y. Wang *et al.*, "Carbon peak and carbon neutrality in China: Goals, implementation path and prospects," *China Geology*, vol. 4, no. 4, pp. 720-746, 2021. https://doi.org/10.31035/cg2021083

[2] J. M. P. Pérez, F. P. G. Márquez, and D. R. Hernández, "Economic viability analysis for icing blades detection in wind turbines," *Journal of Cleaner Production*, vol. 135, pp. 1150-1160, 2016. https://doi.org/10.1016/j.jclepro.2016.07.026

[3] Y. Liu, H. Cheng, X. Kong, Q. Wang, and H. Cui, "Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning," *Energy Science & Engineering*, vol. 7, no. 6, pp. 2633-2645, 2019. https://doi.org/10.1002/ese3.449

[4] Y. Sirui, S. Mengjie, G. Runmiao, B. Jiwoong, Z. Xuan, and Z. Shiqiang, "A review of icing prediction techniques for four typical surfaces in low-temperature natural environments," *Applied Thermal Engineering*, vol. 241, p. 122418, 2024. https://doi.org/10.1016/j.applthermaleng.2024.122418

[5] W. Wang, Y. Xue, C. He, and Y. Zhao, "Review of the typical damage and damage-detection methods of large wind turbine blades," *Energies*, vol. 15, no. 15, p. 5672, 2022. https://doi.org/10.3390/en15155672

[6] X. Dong, D. Gao, J. Li, Z. Jincao, and K. Zheng, "Blades icing identification model of wind turbines based on SCADA data," *Renewable Energy*, vol. 162, pp. 575-586, 2020. https://doi.org/10.1016/j.renene.2020.07.049

[7] I. Kabardin *et al.*, "Optical methods for measuring icing of wind turbine blades," *Energies*, vol. 14, no. 20, p. 6485, 2021. https://doi.org/10.3390/en14206485

[8] H. Lee, Y. M. Hwang, J. Lee, N.-W. Kim, and S.-K. Ko, "A drone-driven X-ray image-based diagnosis of wind turbine blades for reliable operation of wind turbine," *IEEE Access*, 2024. https://doi.org/10.1109/ACCESS.2024.3388494

[9] V. Daniliuk, Y. Xu, R. Liu, T. He, and X. Wang, "Ultrasonic de-icing of wind turbine blades: Performance comparison of perspective transducers," *Renewable Energy*, vol. 145, pp. 2005-2018, 2020. https://doi.org/10.1016/j.renene.2019.07.102

[10] P. Zhu, X. Feng, Z. Liu, M. Huang, H. Xie, and M. A. Soto, "Reliable packaging of optical fiber Bragg grating sensors for carbon fiber composite wind turbine blades," *Composites Science and Technology*, vol. 213, p. 108933, 2021. https://doi.org/10.1016/j.compscitech.2021.108933

[11] P. Rizk, "Hyper spectral imaging applied for the detection of wind turbine blade damage and icing," *Remote Sensing Applications: Society and Environment*, vol. 18, p. 100291, 2020. https://doi.org/10.1016/j.rsase.2020.100291

[12] C. Q. G. Muñoz, F. P. G. Márquez, and J. M. S. Tomás, "Ice detection using thermal infrared radiometry on wind turbine blades," *Measurement*, vol. 93, pp. 157-163, 2016. https://doi.org/10.1016/j.measurement.2016.06.064

[13] P. Guo and D. Infield, "Wind turbine blade icing detection with multi-model collaborative monitoring method," *Renewable Energy*, vol. 179, pp. 1098-1105, 2021. https://doi.org/10.1016/j.renene.2021.07.120

[14] R. Yue, "Spatio-temporal feature alignment transfer learning for cross-turbine blade icing detection of wind turbines," *IEEE Transactions on Instrumentation and Measurement*, 2024. https://doi.org/10.1109/TIM.2024.3350147

[15] M. Kreutz, "Machine learning-based icing prediction on wind turbines," *Procedia Cirp*, vol. 81, pp. 423-428, 2019. https://doi.org/10.1016/j.procir.2019.03.073

[16] X. Yang, T. Ye, Q. Wang, and Z. Tao, "Diagnosis of blade icing using multiple intelligent algorithms," *Energies*, vol. 13, no. 11, p. 2975, 2020. https://doi.org/10.3390/en13112975

[17] J. Xu, W. Tan, and T. Li, "Predicting fan blade icing by using particle swarm optimization and support vector machine algorithm," *Computers & Electrical Engineering*, vol. 87, p. 106751, 2020. https://doi.org/10.1016/j.compeleceng.2020.106751

[18] H. Meng, X. Huang, J. Liu, and Z. Han, "Forecast of fan blade icing combing with random forest and SVM," *Electr. Meas. Instrum*, vol. 57, pp. 66-71, 2020.

[19] T. Tao *et al.*, "Wind turbine blade icing diagnosis using hybrid features and Stacked-XGBoost algorithm," *Renewable Energy*, vol. 180, pp. 1004-1013, 2021. https://doi.org/10.1016/j.renene.2021.09.008

[20] C. Xiong, "Short-term icing status prediction model of wind turbine blades based on Bi-LSTM and SVM models," *Sichuan Electric Power Technology*, vol. 44, no. 3, pp. 88-94, 2021.

[21] L. Dazhong, "Prediction method of fan blade icing based on deep fully connected neural network," *Electric Power Science and Engineering*, vol. 35, no. 4, p. 39, 2019.

[22] C. Tao, "Wind turbine blade icing prediction using focal loss function and cnn-attention-gru algorithm," *Energies*, vol. 16, no. 15, p. 5621, 2023. https://doi.org/10.3390/en16155621

[23] G. Jiang, "Imbalanced learning for wind turbine blade icing detection via spatio-temporal attention model with a self-adaptive weight loss function," *Expert Systems with Applications*, vol. 229, p. 120428, 2023. https://doi.org/10.1016/j.eswa.2023.120428

[24] L. Ying, "Graph temporal attention network for imbalanced wind turbine blade icing prediction," *IEEE Sensors Journal*, 2024. https://doi.org/10.1109/JSEN.2024.3358873

[25] W. Tian, "A multilevel convolutional recurrent neural network for blade icing detection of wind turbine," *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20311-20323, 2021. https://doi.org/10.1109/JSEN.2021.3093726

[26] X. Wang, Z. Zheng, G. Jiang, Q. He, and P. Xie, "Detecting wind turbine blade icing with a multiscale long short-term memory network," *Energies*, vol. 15, no. 8, p. 2864, 2022. https://doi.org/10.3390/en15082864

[27] J. Xiao, C. Li, B. Liu, J. Huang, and L. Xie, "Prediction of wind turbine blade icing fault based on selective deep ensemble model," *Knowledge-based systems*, vol. 242, p. 108290, 2022. https://doi.org/10.1016/j.knosys.2022.108290

[28] R. Tong, P. Li, L. Gao, X. Lang, A. Miao, and X. Shen, "A novel ellipsoidal semisupervised extreme learning machine algorithm and its application in wind turbine blade icing fault detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-16, 2022. https://doi.org/10.1109/TIM.2022.3205920

[29] X. Cheng, F. Shi, X. Liu, M. Zhao, and S. Chen, "A novel deep class-imbalanced semisupervised model for wind turbine blade icing detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2558-2570, 2021. https://doi.org/10.1109/TNNLS.2021.3102514

[30] W. Chen, L. Cheng, Z. Chang, B. Wen, and P. Li, "Wind turbine blade icing detection using a novel bidirectional gated recurrent unit with temporal pattern attention and improved coot optimization algorithm," *Measurement Science and Technology*, vol. 34, no. 1, p. 014004, 2022. https://doi.org/10.1088/1361-6501/ac8db1

[31] X. Cheng, F. Shi, M. Zhao, G. Li, H. Zhang, and S. Chen, "Temporal attention convolutional neural network for estimation of icing probability on wind turbine blades," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6371-6380, 2021. https://doi.org/10.1109/TIE.2021.3090702

[32] C. Tao, T. Tao, S. He, X. Bai, and Y. Liu, "Wind turbine blade icing diagnosis using B-SMOTE-Bi-GRU and RFE combined with icing mechanism," *Renewable Energy*, vol. 221, p. 119741, 2024. https://doi.org/10.1016/j.renene.2023.119741

[33] X. Liu, J. Liu, J. Liu, Y. Zhao, Z. Yang, and T. Ding, "A bayesian deep learning-based wind power prediction model considering the whole process of blade icing and De-icing," *IEEE Transactions on Industrial Informatics*, 2024. https://doi.org/10.1109/TII.2024.3379668

[34] K. Yin and R. Jesse, "Attention is all you sign: Sign language translation with transformers, sign language recognition," *Translation and Production (SLRTP) Workshop-Extended Abstracts*, 2020.

[35] Y. SHI, W. Yu, and W. Shui-qing, "Machine translation system based on self-Attention model," *Computer and Modernization*, vol. 7, no. 9, 2019. https://doi.org/10.3969/j.issn.1006-2475.2019.07.002

[36] A. Sasinthiran, G. Sakthivel, and R. Ragala, "A review of artificial intelligence applications in wind turbine health monitoring," *International Journal of Sustainable Energy*, vol. 43, no. 1, p. 2326296, 2024.

[37] C. Yuan, F. Wang, M. Jiang, P. Long, S. Yu, and Y. Liu, "Wavelet FCNN: A deep time series classification model for wind turbine blade icing detection," *ar Xiv preprint, ar Xiv:1902.05625*, 2019.

[38] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," presented at the In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 1578-1585). IEEE, 2017.

[39] L. Xiong, C. Yi, Q. Xiong, and S. Jiang, "SEA-NET: Medical image segmentation network based on spiral squeeze-and-excitation and attention modules," *BMC Medical Imaging*, vol. 24, no. 1, p. 17, 2024. https://doi.org/10.1186/s12880-024-01194-8

[40] A. F. Gad, "Pygad: An intuitive genetic algorithm python library," *Multimedia Tools and Applications*, vol. 83, no. 20, pp. 58029-58042, 2024. https://doi.org/10.1007/s11042-023-17167-y

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312

[41] J. Águila-León, C. Vargas-Salgado, D. Díaz-Bello, and C. Montagud-Montalvá, "Optimizing photovoltaic systems: a meta-optimization approach with GWO-Enhanced PSO algorithm for improving MPPT controllers," *Renewable Energy*, vol. 230, p. 120892, 2024. https://doi.org/10.1016/j.renene.2024.120892

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6: 2035-2054, 2025
DOI: 10.55214/25768484.v9i6.8312