

Article

Wind Turbine Blade Icing Prediction Using Focal Loss Function and CNN-Attention-GRU Algorithm

Cheng Tao ¹, Tao Tao ² , Xinjian Bai ¹ and Yongqian Liu ^{1,*}

¹ State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (NCEPU), School of New Energy, North China Electric Power University, Beijing 102206, China

² China Southern Power Grid Technology Co., Ltd., Guangzhou 510080, China

* Correspondence: yqliu@ncepu.edu.cn

Abstract: Blade icing seriously affects wind turbines' aerodynamic performance and output power. Timely and accurately predicting blade icing status is crucial to improving the economy and safety of wind farms. However, existing blade icing prediction methods cannot effectively solve the problems of unbalanced icing/non-icing data and low prediction accuracy. In order to solve the above problems, this paper proposes a wind turbine blade icing prediction method based on the focal loss function and CNN-Attention-GRU. First, the recursive feature elimination method combined with the physical mechanism of icing is used to extract features highly correlated with blade icing, and a new feature subset is formed through a sliding window algorithm. Then, the focal loss function is utilized to assign more weight to the ice samples with a lower proportion, addressing the significant class imbalance between the ice and non-ice categories. Finally, based on the CNN-Attention-GRU algorithm, a blade icing prediction model is established using continuous 24-h historical data as the input and the icing status of the next 24 h as the output. The model is compared with advanced neural network models. The results show that the proposed method improves the prediction accuracy and F_1 score by an average of 6.41% and 4.27%, respectively, demonstrating the accuracy and effectiveness of the proposed method.

Keywords: wind turbine blade; icing prediction; SCADA; focal loss; neural networks



Citation: Tao, C.; Tao, T.; Bai, X.; Liu, Y. Wind Turbine Blade Icing Prediction Using Focal Loss Function and CNN-Attention-GRU Algorithm. *Energies* **2023**, *16*, 5621. <https://doi.org/10.3390/en16155621>

Academic Editors: Frede Blaabjerg and Dao Zhou

Received: 18 June 2023

Revised: 14 July 2023

Accepted: 17 July 2023

Published: 27 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blade icing refers to the phenomenon where moisture on the blade's surface condenses and forms ice when the ambient temperature is below freezing [1,2]. Blade icing is a common issue when wind turbines operate in cold environments. The accumulation of ice can affect the aerodynamic performance of the blades, leading to a decrease in power generation efficiency or even shutdown [3–5]. Furthermore, blade icing can cause an uneven load distribution and irregular vibrations, which further impact the reliability and lifespan of the wind turbine [6,7]. Therefore, the accuracy and timeliness of blade icing prediction are crucial for the operation and management of wind farms. The accurate prediction of icing can help operators take appropriate measures, such as activating heating systems and cleaning the blades, minimizing the negative impacts of icing [8–10]. This helps improve power generation efficiency, reduces maintenance costs, and ensures the stable operation of wind turbines.

At present, blade icing prediction methods can be broadly classified into direct methods and indirect methods. Direct methods involve installing sensors around the blades to analyze changes in monitored variables caused by ice accumulation, such as variations in capacitance and resistance [11], variations in ultrasound signals [12], variations in thermal infrared radiation signals [13], changes in hyperspectral images [14], rotor speed and pitch angle variations [15], power variations [16], etc. However, direct methods rely on precise and reliable sensors, and most focus on diagnosing icing conditions rather than predicting

them in advance. They cannot provide early warnings about icing conditions. On the other hand, indirect methods involve establishing icing prediction models to forecast blade icing status in advance. These methods can provide accurate icing status predictions, enabling operators to receive timely warnings. As a result, more and more researchers are adopting indirect methods to study blade icing prediction.

Indirect methods can be further divided into physics-based methods and data-driven methods. Physics-based methods primarily study the relationship between meteorological conditions or the operational parameters of wind turbines and icing, and construct physical models for blade icing [17]. This method requires significant professional knowledge, which is challenging, complex, and difficult to model. On the other hand, data-driven methods do not rely on complex theoretical models. They can be applied to various domains and adaptively update the models as the data change. Data-driven methods can also uncover underlying patterns in the data and perform well in tasks such as prediction, classification, and clustering. Currently, many researchers employ data-driven methods to predict the blade icing status of wind turbines, which mainly involves feature extraction and model construction.

In terms of feature extraction, Kreutz et al. [18] utilized their empirical knowledge to select environmental temperature and wind speed at the hub height from SCADA data, as well as the humidity, pressure, ground temperature, temperature at 200 m height, ground wind speed, and wind speed at 200 m height from meteorological forecast data as features for blade icing prediction. Xiao et al. [19] ranked the importance of 26 features in SCADA data using the chi-square test and recursively removed features with low importance to obtain the optimal feature set. Peng et al. [20] employed dynamic principal component analysis for feature extraction and ultimately selected yaw angle, nacelle temperature, and generator speed as the features for blade icing. Ma et al. [21] constructed a severity index for the blade icing and selected wind direction, environmental temperature, pitch angle, power, and wind speed as features based on statistical characteristics and the trend of icing data over continuous periods.

Regarding model construction, Bai et al. [22] developed an RFECV-TSVM icing diagnosis model, which addressed the issue of class imbalance in blade icing data by generating pseudo-samples. However, this method is not suitable for icing prediction. Tao et al. [23] used random under-sampling to address the class imbalance issue in icing data and established a Stacked-XGBoost blade icing diagnosis model. Xiao et al. [19] proposed the Selective Deep Ensemble model based on the Group Method of Data Processing as a blade icing prediction model. Li et al. [24] utilized backpropagation neural networks and radial basis function neural networks as blade icing prediction models and evaluated the relative percentage errors of both models. Peng et al. [20] constructed a blade icing prediction model based on the backpropagation self-organizing clustering algorithm. Kreutz et al. [25] employed a five-layer convolutional neural network as a blade icing prediction model to forecast wind turbine blades' icing conditions in the next 24 h. Ma et al. [21] developed a four-layer deep belief network as a blade icing prediction model.

In summary, the research on wind turbine blade icing prediction faces the following challenges:

(1) There is a significant class imbalance issue in blade icing data. Most of the time, wind turbine blades are in a normal state, resulting in a low proportion of icing data. This imbalance makes it easy for machine learning models to overlook the minority of icing instances during training, increasing the difficulty of wind turbine blade icing prediction.

(2) The accuracy of wind turbine blade icing prediction models is low. Existing methods often use a single neural network module to construct blade icing prediction models, aiming to improve the prediction performance by increasing the number of layers, resulting in a model that performs well on the training set but poorly on unseen data.

To address the abovementioned issues, this paper proposes a wind turbine blade icing prediction method based on the focal loss function and CNN-Attention-GRU algorithm. The main contributions of this paper are as follows:

(1) The focal loss function is employed as the model's loss function to address the class imbalance issue. By assigning more weight to the icing data, the model focuses more on the minority class samples, enhancing its learning ability for icing information.

(2) A wind turbine blade icing prediction model based on CNN-Attention-GRU is established. This can better mine the potential icing information in the SCADA historical data. Compared to previous methods, the proposed model achieves a higher prediction accuracy.

The remaining parts of the paper are as follows: Section 2 describes the data preprocessing process, including feature extraction and dataset construction. Section 3 introduces the wind turbine blade icing prediction model based on the focal loss function and CNN-Attention-GRU. Section 4 introduces the dataset and presents case studies and analyses. Section 5 concludes the paper.

2. Data Processing

2.1. Feature Extraction

2.1.1. Recursive Feature Elimination

Recursive Feature Elimination (RFE) is a feature extraction method to select a subset of the most predictive features for modelling tasks. This method works by recursively training a model and eliminating relatively unimportant features [26]. The basic principle of RFE is as follows:

First, a model is trained based on the original feature set, and the importance scores of each feature are calculated. The importance scores of features can be calculated using different models and feature evaluation methods (such as decision trees, linear regression, random forests, etc.). Second, based on the importance scores of features, a subset of features is selected, consisting of the top-ranked features in terms of importance for the current iteration.

Then, this feature subset is used to train a model, and the performance metrics of the model (such as accuracy) are calculated. If the performance metrics of the selected feature subset meet a predetermined threshold or a predetermined number of features have been eliminated, the process is stopped. Otherwise, the process returns to the second step, continuing to iterate and select feature subsets until the stopping criteria are met.

The advantages of the Recursive Feature Elimination (RFE) method are that it can automatically select a subset of features, making the final model more simplified and interpretable. Gradually eliminating unimportant features reduces the risk of overfitting and improves the model's generalization ability. Additionally, RFE can rank the importance of features, facilitating a better understanding of the data and the modelling process.

In this paper, RFE was applied with a random forest as the estimator, and the following 10 features were identified as necessary for the dataset: actual power, wind speed, generator speed, environmental temperature, cabin temperature, yaw angle, wind direction, gearbox oil temperature, gearbox bearing temperature, and generator temperature.

2.1.2. Feature Construction Based on Icing Physics

Tao et al. [23], based on the icing physics process and the analysis of ice formation mechanisms, constructed diagnostic features for blade icing. Based on this, this paper selects four parameters: theoretical power, blade tip speed ratio, wind speed squared, and wind speed cubed, as features extracted based on empirical knowledge.

(1) Theoretical power

Theoretical power is an essential feature for predicting blade icing in wind turbines. It describes the theoretical output power curve of the turbine at different wind speeds and is calculated using the following formula.

$$P_{ideal} = f(P_{actual}) \quad (1)$$

where P_{ideal} is the theoretical power; $f(\cdot)$ is the functional relationship between theoretical power and actual power; and P_{actual} is the actual power.

However, the original SCADA data do not directly provide the theoretical power. Therefore, we need to fit and obtain it through the following steps. First, we need to remove invalid data, which includes data points where the wind speed is less than 3 m/s but the actual power is non-zero, data points where the wind speed is greater than 3 m/s but the actual power is zero, and data points where the wind speed is greater than 25 m/s. Next, we sort the wind speeds in ascending order and divide them into 100 equally spaced intervals. Using the quartile method, we divide the actual power into four equal parts within each interval and obtain the first quartile (25th percentile) and the third quartile (75th percentile) values. We obtain the interquartile range by calculating the difference between these two values. Finally, we can calculate the range of normal samples within each interval using the formula, obtaining the fitted theoretical power.

$$[I_b, I_e] = [Q_1 - 1.5I_{QR}, Q_3 + 1.5I_{QR}] \quad (2)$$

where $[I_b, I_e]$ is the range of normal samples in each interval calculated by the quartile method; Q_1 is the first quartile; I_{QR} is the interquartile range; and Q_3 is the third quartile.

Exclude the outliers outside the $[I_b, I_e]$ intervals and calculate the average wind speed and actual power within each interval to fit the theoretical power curve.

(2) Tip speed ratio

$$\lambda = \frac{\omega R}{v} \quad (3)$$

where λ is the tip speed ratio; ω is the rotor's rotation speed; R is the radius of the blade's rotation plane; and v is the wind speed.

(3) The square of wind speed

$$v' = v^2 \quad (4)$$

(4) The cube of wind speed

$$v'' = v^3 \quad (5)$$

This paper used a combination of RFE and icing physics knowledge for feature extraction. The final selected features from the dataset are shown in Table 1:

Table 1. Final features for blade icing prediction.

Feature Names	
Actual power	Gear oil temperature
Wind speed	Gearbox bearing temperature
Generator speed	Generator temperature
Ambient temperature	Theoretical power
Nacelle temperature	Tip speed ratio
Yaw angle	The square of wind speed
Wind direction	The cube of wind speed

RFE removes the least important features one at a time and extracts the features with a high correlation to icing, achieving the compression of feature space dimension, which greatly reduces the model computation time while ensuring the prediction accuracy. The features extracted based on icing physics enhance the non-linear mapping ability of the model and increase the accuracy of icing prediction while compensating for the shortcomings of machine learning such as lack of interpretability. Therefore, the combination of

machine learning methods and icing physics for feature extraction is of great significance for icing prediction.

2.2. Constructing Dataset Based on Sliding Window Algorithm

The sliding window algorithm is commonly used for processing subsequences or subarrays in sequence data or arrays. Its basic idea is to transform the problem into operations on a sliding window. Using a fixed-size window and defining the starting position, ending position, and sliding step, the algorithm slides the window over the sequence to gradually process the data [27]. It can calculate statistical information, find the maximum or minimum value, and perform other operations within the window. The advantage of the sliding window algorithm is that it can solve the problem in a single pass of the data without the need for multiple traversals of the entire sequence. This makes the sliding window algorithm highly efficient when dealing with large-scale or real-time data. The process of constructing a new dataset using the sliding window is illustrated in Figure 1.

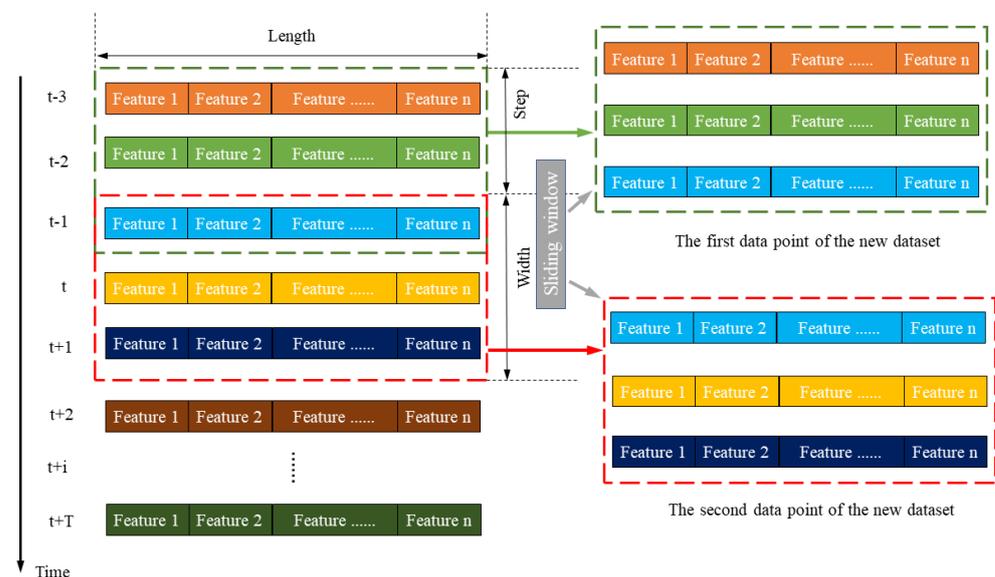


Figure 1. Process of sliding window algorithm.

In Figure 1, the horizontal represents the features and vertical represents the time, where the length = 14 indicates the number of features finally selected. Step = 1 indicates the size of each downward slide of the window. Width = 144 indicates that each window contains 24 h of SCADA data information. Finally, the original two-dimensional data are transformed into a three-dimensional input suitable for the neural network models.

2.3. Max–Min Normalization

Max–Min normalization, also known as interval scaling, is a commonly used method for data normalization. It transforms data into a specified range, typically between [0, 1] or [−1, 1]. Max–Min normalization achieves this by performing a linear transformation on the original data, mapping the minimum value to the minimum value of the target range and the maximum value to the maximum value of the target range while maintaining the relative relationship of other values within this range. It is commonly used in the preprocessing stage of machine learning algorithms to scale the values of different features to the same range, preventing certain features from disproportionately impacting the model. The formula is as follows:

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (6)$$

where x_i are data of each dimension in the dataset before normalization; $\max(x_i)$ is the maximum value of each dimension in the dataset; and $\min(x_i)$ is the minimum value of each dimension in the dataset.

3. Model Building

3.1. Focal Loss Function

In traditional cross-entropy loss functions, for imbalanced datasets, the number of samples in the majority class far exceeds the number of samples in the minority class, causing the model to be more biased towards predicting the majority class samples and ignoring the minority class samples [28]. The focal loss function is a type of loss function designed to address the issue of class imbalance. It aims to tackle the difficulty of effectively learning and classifying minority class samples when there are many majority class samples in the data. Focal loss introduces an adjustable balancing parameter to modify the weight relationship between the majority and minority class samples [29].

Specifically, the focal loss function introduces an adjustable parameter called the focusing factor, which adjusts the loss weight for each sample. The focusing factor is calculated based on the predicted probability of each sample. For the majority class samples, the focusing factor is small, reducing their weight. In contrast, for minority class samples, the focusing factor is significant, increasing their weight. This allows the model to pay more attention to the minority class samples, thus improving the issue of class imbalance. The formula for the focal loss functions as follows:

$$FL(p_t) = -(1 - p_t)^\gamma \log(p_t) \quad (7)$$

where p_t is the predicted probability of the sample; and γ is the focusing factor. When $\gamma = 0$, the focal loss function degenerates into the standard cross-entropy loss function. Xiao et al. found that the model achieved the highest accuracy when $\gamma = 2$ in the focal loss function [19]. Therefore, this paper selects $\gamma = 2$.

3.2. GRU Neural Network

Gated Recurrent Unit (GRU) is a Recurrent Neural Networks (RNN) variant used to process and model sequential data. Similar to Long Short-Term Memory (LSTM), GRU introduces a particular type of memory unit to address the issue of long-term dependencies. It uses gate mechanisms to control the flow of information, thereby improving the accuracy and performance of the model [30]. GRU also simplifies the LSTM network structure to some extent while still effectively capturing long-term dependencies in sequences [31]. Since icing prediction is a time series forecasting problem and the data contain significant temporal information related to icing, GRU is well-suited for handling sequential data with long-term dependencies. It can effectively capture and retain long-term information in the sequences. Therefore, this paper chooses the GRU model.

The core component of the GRU model is the GRU unit. Each GRU unit consists of two essential parts: the reset and update gates. The GRU model controls the flow of information and retention through the mechanisms of the reset and update gates, enabling sequential data modelling. The reset gate controls the influence of the previous state on the current input. The update gate determines how much new information should be merged into the previous state. Compared to the LSTM model, the GRU model has fewer parameters, higher computational efficiency, and is suitable for medium-scale sequential modelling tasks. The schematic diagram of the GRU unit is shown in Figure 2.

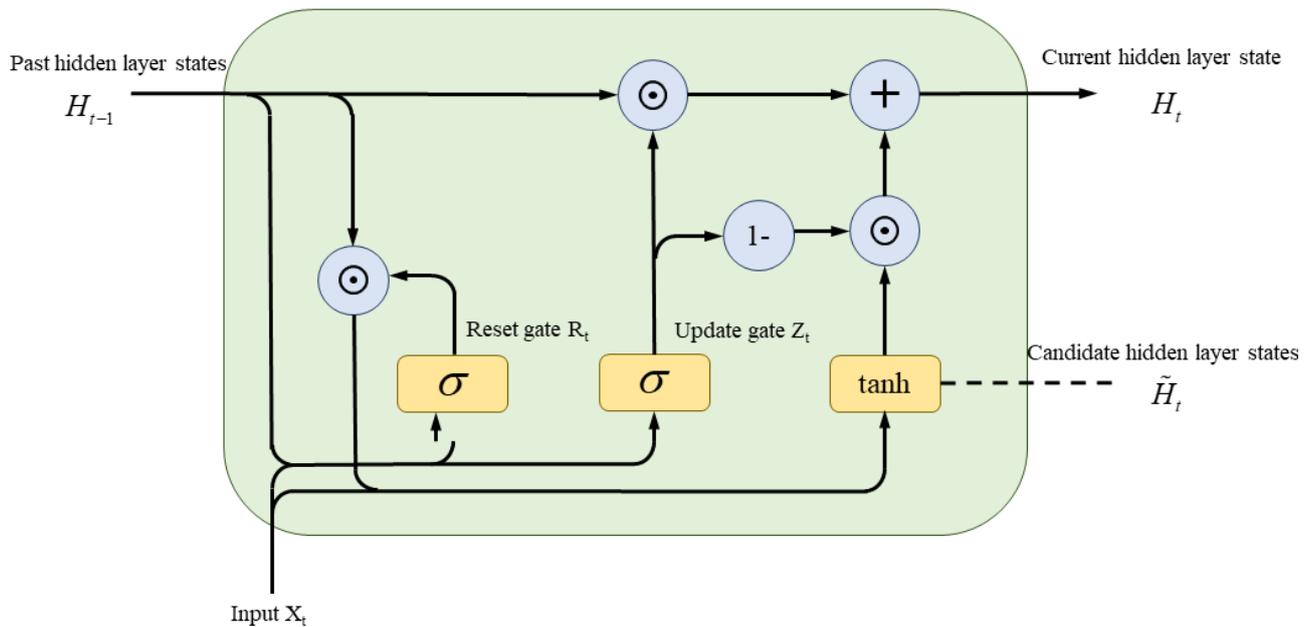


Figure 2. The structure of the GRU unit.

3.3. CNN Neural Network

The Convolutional Neural Network (CNN) is a type of deep learning model widely used in computer vision tasks, particularly image recognition and classification tasks. CNNs extract features from the image data through convolutional and pooling operations and perform classification or regression through fully connected layers [32].

The core components of a CNN model include the convolutional layer and the pooling layer. The convolutional layer applies convolutional operations between the original image and a set of learnable filters to extract local features from the image. The convolution operation involves element-wise multiplication and summation between the filters and the input image, resulting in feature maps. The convolutional layer captures different image features, such as edges and textures, by using multiple filters. The pooling layer reduces the spatial dimensions of the feature maps while preserving the essential features. Common pooling operations include max pooling and average pooling. The pooling layer divides the feature maps into non-overlapping regions and selects the maximum value or computes the average value within each region. This helps to reduce computational complexity and the number of parameters while improving the robustness and generalization ability of the model.

3.4. Attention Mechanism

Attention is a technique to enhance the model's focus on different input parts. It is commonly used in natural language processing (NLP) and computer vision (CV) tasks. It dynamically assigns different weights to different input elements, allowing the model to concentrate on important information relevant to the current task during the processing.

The main goal of the attention is to address the performance degradation issue in models when dealing with long sequences or significant inputs, especially in tasks involving long-term dependencies. By introducing the attention, the model associates a weight with each element in a sequence, representing the model's level of focus on that element. By calculating the similarity between elements and converting it into weights, different levels of attention can be assigned to different positions in the sequence, allowing the model to pay more attention to important information during the processing [33].

The CNN-Attention-GRU model structure used in this study for blade icing prediction is shown in Figure 3. This paper utilizes SCADA historical data from the past 24 h to predict the blade icing status for the next 24 h. The model first extracts local icing features

from the input data through convolutional layers and pooling layers. Then, it applies the attention to aggregate these features with different weights. Finally, the GRU layer is used to model and process the temporal data, predicting the final results of blade icing.

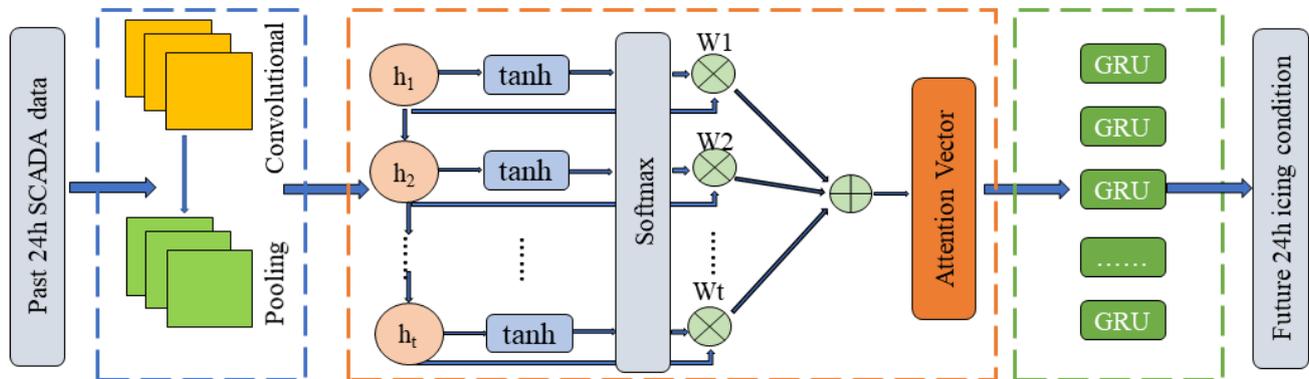


Figure 3. The schematic diagram of the CNN-Attention-GRU model structure.

3.5. Evaluation Metrics

The confusion matrix is a commonly used metric in machine learning for classification problems. When applying the confusion matrix to blade icing prediction, the specific meanings of each indicator are as follows: True Positive (TP) represents the number of times the model correctly predicts the normal state data as the normal state data. True Negative (TN) represents the number of times the model correctly predicts the icing state data as the icing state data. False Positive (FP) represents the number of times the model incorrectly predicts the icing state data as normal. False Negative (FN) represents the number of times the model incorrectly predicts the normal state data as the icing state data. Based on the above information, the evaluation metrics used in this study are calculated using the following formulas:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

where *Accuracy* refers to the accuracy rate; *Precision* refers to the precision rate; *Recall* refers to the recall rate of normal samples; F_1 refers to the harmonic mean of *Precision* and *Recall*. Since *Precision*, *Recall*, and F_1 have redundancy, this paper only uses *Accuracy* and F_1 as the evaluation metrics.

4. Case Study

4.1. Data Description

The dataset used in this study is the SCADA data of wind turbine icing in a mountainous wind farm in Yunnan Province, China. The dataset covers the icing data from five doubly fed asynchronous wind turbine units with the following identifiers: A1, A2, A3, A4, and A5. The data span from 24 January 2018 to 31 December 2018, with a temporal resolution of 10 min. The dataset consists of 28 continuous numerical variables, among which 19 are feature variables. The names and descriptions of the feature variables are presented in Table 2. Table 2 shows the features in the original SCADA data, not the final

features selected after feature extraction. Each sample in the dataset is labelled with a status tag, where icing state data are labelled as “0” and normal state data are labelled as “1”. Icing events at the wind farm occurred only from the night of December 29th to the early morning of December 31st. Therefore, the data selected for analysis include the period from December 24th to December 31st, with data before December 24th not used for training and testing. The quantity and proportion of data for the five wind turbine units are shown in Table 3.

Table 2. Feature parameter names and descriptions for the dataset.

Feature Name	Feature Description	Feature Name	Feature Description
WIND_SPEED	Wind speed	GENGNTMP	Generator temperature
REAL_POWER	The active power of grid-side	GENAPHSA	Current of A-phase
CONVERTER_MOTOR_SPEED	Generator speed	GENAPHSB	Current of B-phase
ROTOR_SPEED	Blade rotation speed	GENAPHSC	Current of C-phase
WIND_DIRECTION	Wind direction	GENVPHSA	Voltage of A-phase
TURYAWDIR	Yaw angle	GENVPHSB	Voltage of B-phase
GBXOILTMP	Temperature of gear oil	GENVPHSC	Voltage of C-phase
GBXSHFTMP	Temperature of gearbox bearing	GENHZ	Frequency of motor
EXLTMP	Temperature of environment	TURPWRREACT	Reactive power
TURINTTMP	Temperature of nacelle		

Table 3. Statistical parameters of the dataset.

Wind Turbine Number	Total Number of Data	Normal Data	Icing Data
A1	1009	857 (84.94%)	152 (15.06%)
A2	1009	857 (84.94%)	152 (15.06%)
A3	997	831 (83.35%)	167 (16.65%)
A4	1009	849 (84.14%)	160 (15.86%)
A5	1009	826 (81.86%)	183 (18.14%)

This paper combines the SCADA data from the same wind farm’s five wind turbines. This increases the information on blade icing in the data, but the ratio of icing to non-icing data is the same as before the combination. The data from units A1, A2, A3, A4, and A5 are used as individual test sets, while the data from the remaining four units are used as training and validation sets. This forms five feature subsets, denoted as A1T, A2T, A3T, A4T, and A5T, as shown in Figure 4.



Figure 4. Schematic diagram of the dataset combination.

The overall process is illustrated in Figure 5. First, the RFE method combined with the physical mechanism of icing is used to extract features highly correlated with blade icing, and a new feature subset is formed through a sliding window algorithm. Kreutz et al. [9] split data into a training dataset, a validation dataset, and a test dataset with a 60/20/20%

ratio, with a high accuracy of model accuracy. However, due to the small amount of data in this paper, the proportion of the training set is adjusted to 70%, the validation set is 10%, and the test set is still 20%. Then, the focal loss function is utilized to assign more weight to the ice samples with a lower proportion, addressing the significant class imbalance between ice and non-ice categories. Finally, based on the CNN-Attention-GRU algorithm, a blade icing prediction model is established using continuous 24-h historical data as the input and the icing status of the next 24 h as the output.

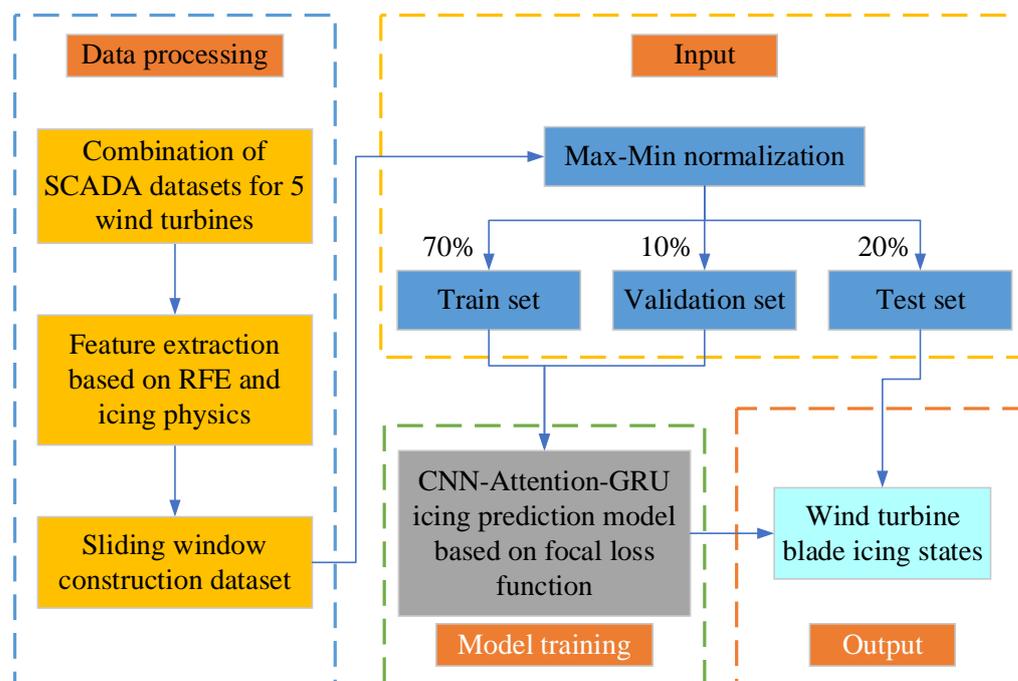


Figure 5. The process of wind turbine blade icing prediction.

4.2. Verification of the Validity of the Focal Loss Function

Four neural network models were trained using binary a cross-entropy loss for icing prediction to further validate the effectiveness of the focal loss function proposed in this study for addressing the class imbalance issue in wind turbine blade icing. The results were compared with the accuracy and F_1 scores obtained using different loss functions for the same models, as shown in Figures 6 and 7. After applying the focal loss function, all algorithms showed a noticeable improvement in prediction accuracy. The average accuracy across the four algorithms increased by 4.08%, with the CNN-LSTM model exhibiting the most significant improvement of 9.68% and the CNN-Attention-GRU model showing the minor improvement of 0.36%. Similarly, the average F_1 improved by 3.48%, with the CNN-LSTM model experiencing the most enormous improvement of 9.81% and the CNN-Attention-GRU model showing the smallest improvement of 0.07%. These results indicate that the focal loss function proposed in this study is suitable for addressing the class imbalance issue in wind turbine blade icing. The minimal improvement in accuracy for the CNN-Attention-GRU model also suggests its stability and suitability for icing prediction tasks.

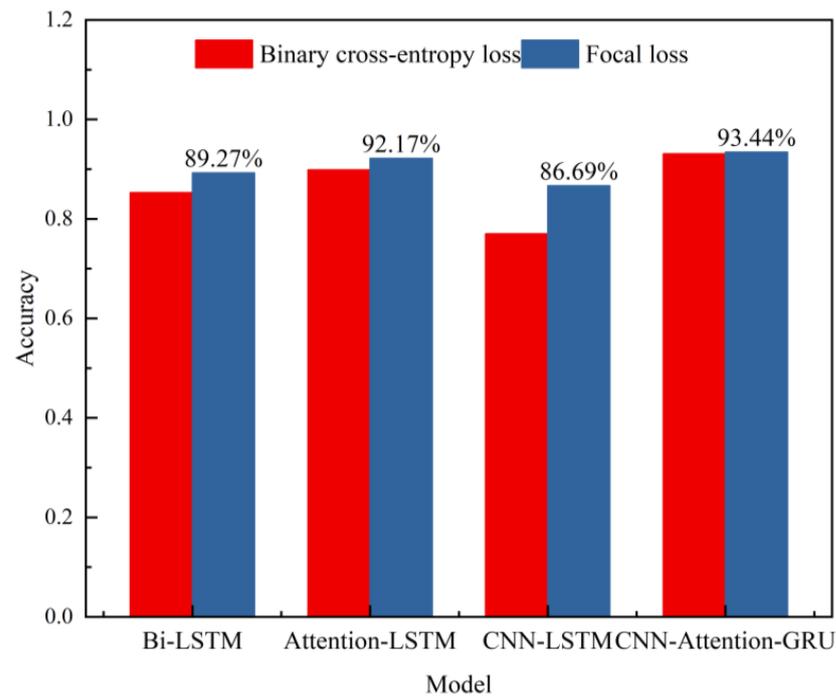


Figure 6. Prediction accuracy of different loss functions in various models.

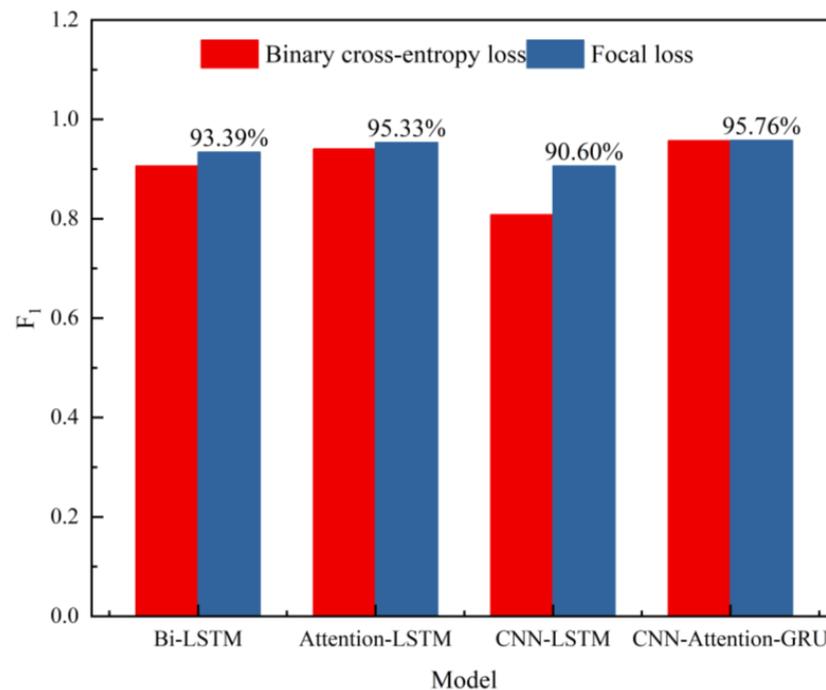


Figure 7. Prediction F_1 of different loss functions in various models.

4.3. Prediction Accuracy Validation of the CNN-Attention-GRU Model

The dataset was divided into 70% for training, 10% for validation, and 20% for testing. The focal loss function was used, and the CNN-Attention-GRU neural network model was employed to predict the icing conditions of wind turbines for the next 24 h compared with 10 other neural network models and the results of the test set are shown in Table 4. Where, the bolded portion represents the highest prediction accuracy for each model in the same dataset.

Table 4. Icing prediction results of different datasets in various models.

Model	Accuracy						F ₁					
	A1T	A2T	A3T	A4T	A5T	Average	A1T	A2T	A3T	A4T	A5T	Average
LSTM	0.9222	0.9597	0.8787	0.9694	0.6667	0.8793	0.9482	0.9739	0.9266	0.9808	0.7425	0.9144
GRU	0.9083	0.7222	0.8082	0.9708	0.8403	0.8500	0.9384	0.8279	0.8887	0.9816	0.898	0.9069
Bi-LSTM	0.9653	0.7806	0.9859	0.9847	0.7472	0.8927	0.9775	0.8556	0.9909	0.9903	0.8553	0.9339
Bi-GRU	0.975	0.8458	0.7884	0.9736	0.7472	0.8660	0.9839	0.9095	0.8786	0.9833	0.8553	0.9221
Attention-LSTM	0.9708	0.9403	0.9972	0.9333	0.7667	0.9217	0.9814	0.9628	0.9982	0.959	0.865	0.9533
Attention-GRU	0.9819	0.8444	0.9986	0.8556	0.7472	0.8855	0.9884	0.9079	0.9991	0.9152	0.8553	0.9332
CNN-LSTM	0.7597	0.9389	0.976	0.9153	0.7444	0.8669	0.8207	0.9613	0.9846	0.9484	0.8149	0.9060
CNN-GRU	0.6014	0.7722	0.8999	0.9597	0.6069	0.7680	0.6627	0.8316	0.9386	0.9748	0.6705	0.8156
CNN-Flatten	0.9403	0.8181	0.969	0.9625	0.7639	0.8908	0.9609	0.8868	0.9801	0.9765	0.8636	0.9336
CNN-Attention-LSTM	0.9764	0.9028	0.9492	0.8361	0.7472	0.8823	0.9848	0.9383	0.9679	0.9048	0.8553	0.9302
CNN-Attention-GRU	0.9153	0.9278	0.9803	0.9069	0.9417	0.9344	0.9434	0.9521	0.9873	0.9437	0.9615	0.9576

The mixed distribution of data from the five turbines has some influence on the prediction results. The best-performing models on the five test sets are different, and no clear pattern is based solely on these results. To address this, the average accuracy and F₁ of the predictions from the mixed data of the five turbines were calculated, and the results are shown in Figures 8 and 9. The different coloured bars in the figures represent the prediction results of different neural network models on the test set, with the red bar on the far right representing the CNN-Attention-GRU model based on the focal loss function used in this study. Compared to the other 10 models, the average accuracy was improved by 6.41%, with the CNN-GRU model having the highest improvement rate of 16.64%. The average F₁ was improved by 4.27%, with the CNN-GRU model having the highest improvement rate of 14.20%. These results indicate that the model used in this paper performs better than other models in the icing prediction and demonstrates good applicability.

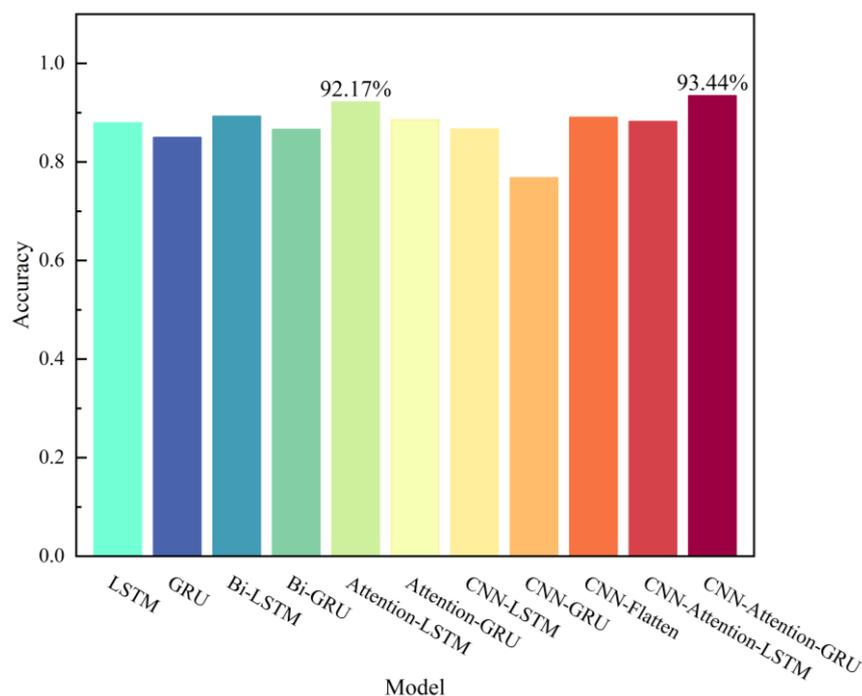


Figure 8. The average prediction accuracy of different models.

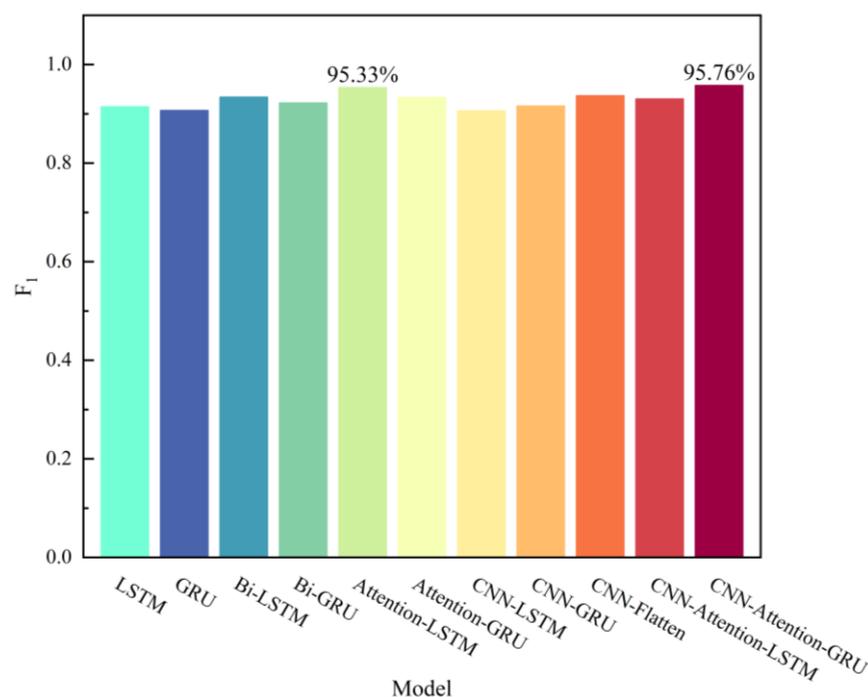


Figure 9. The average prediction F_1 of different models.

5. Conclusions

This study focuses on wind turbine blade icing predictions based on SCADA data from five wind turbines in a wind farm. The research is conducted from two aspects: data processing and model construction. Firstly, a method combining RFE and empirical icing knowledge extract features were highly correlated with blade icing. Then, the data from the five turbines are combined, and a new feature subset is formed using a sliding window algorithm. Finally, the focal loss function is employed to address the issue of class imbalance, and different neural network models are used to evaluate the prediction accuracy of the proposed CNN-Attention-GRU model. The main conclusions of this study are as follows:

The neural network model based on the focal loss function can address the issue of class imbalance in wind turbine blade icing data. Compared to other neural network models that use a binary cross-entropy loss function, this method achieves an average accuracy improvement of 4.08%, with the CNN-LSTM model showing the highest improvement of 9.68%. The average F_1 improves by 3.48%, with the CNN-LSTM model showing the highest improvement of 9.81%.

The CNN-Attention-GRU model can address the issue of low accuracy in wind turbine blade icing prediction. Compared to other neural network models, this method achieves an average accuracy improvement of 6.41%, with the CNN-GRU model showing the highest improvement of 16.64%. The average F_1 improves by 4.27%, with the CNN-GRU model showing the most significant improvement of 14.20%. In conclusion, the proposed method in this paper effectively improves the accuracy of wind turbine blade icing prediction and is more suitable for the problem of blade icing prediction.

Author Contributions: C.T.: conceptualization, methodology, software, writing—original draft. T.T.: writing—review and editing, methodology. X.B.: formal analysis, data curation. Y.L.: supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [the National Key Research and Development Program of China] grant number [No.2019YFE0104800]. And The APC was funded by [Research on smart operation control technologies for offshore wind farms].

Data Availability Statement: Wind farms have confidentiality requirements for this data and we cannot publish the data we have used.

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

References

1. Wang, W.; Qin, C.; Zhang, J.; Wen, C.; Xu, G. Correlation Analysis of Three-Parameter Weibull Distribution Parameters with Wind Energy Characteristics in a Semi-Urban Environment. *Energy Rep.* **2022**, *8*, 8480–8498. [[CrossRef](#)]
2. Liu, H.; Li, Y.; Duan, Z.; Chen, C. A Review on Multi-Objective Optimization Framework in Wind Energy Forecasting Techniques and Applications. *Energy Convers. Manag.* **2020**, *224*, 113324. [[CrossRef](#)]
3. Ibrahim, G.M.; Pope, K.; Naterer, G.F. Extended Scaling Approach for Droplet Flow and Glaze Ice Accretion on a Rotating Wind Turbine Blade. *J. Wind Eng. Ind. Aerodyn.* **2023**, *233*, 105296. [[CrossRef](#)]
4. Dai, Y.; Xie, F.; Li, B.; Wang, C.; Shi, K. Effect of Blade Tips Ice on Vibration Performance of Wind Turbines. *Energy Rep.* **2023**, *9*, 622–629. [[CrossRef](#)]
5. Hu, Q.; Xu, X.; Leng, D.; Shu, L.; Jiang, X.; Virk, M.; Yin, P. A Method for Measuring Ice Thickness of Wind Turbine Blades Based on Edge Detection. *Cold Reg. Sci. Technol.* **2021**, *192*, 103398. [[CrossRef](#)]
6. Jin, J.Y.; Virk, M.S. Experimental Study of Ice Accretion on S826 & S832 Wind Turbine Blade Profiles. *Cold Reg. Sci. Technol.* **2020**, *169*, 102913. [[CrossRef](#)]
7. Hacıfendioğlu, K.; Başağa, H.B.; Yavuz, Z.; Karimi, M.T. Intelligent Ice Detection on Wind Turbine Blades Using Semantic Segmentation and Class Activation Map Approaches Based on Deep Learning Method. *Renew. Energy* **2022**, *182*, 1–16. [[CrossRef](#)]
8. Guk, E.; Son, C.; Rieman, L.; Kim, T. Experimental Study on Ice Intensity and Type Detection for Wind Turbine Blades with Multi-Channel Thermocouple Array Sensor. *Cold Reg. Sci. Technol.* **2021**, *189*, 103297. [[CrossRef](#)]
9. Kreutz, M.; Alla, A.A.; Eisenstadt, A.; Freitag, M.; Thoben, K.-D. Ice Detection on Rotor Blades of Wind Turbines Using RGB Images and Convolutional Neural Networks. *Procedia CIRP* **2020**, *93*, 1292–1297. [[CrossRef](#)]
10. Madi, E.; Pope, K.; Huang, W.; Iqbal, T. A Review of Integrating Ice Detection and Mitigation for Wind Turbine Blades. *Renew. Sustain. Energy Rev.* **2019**, *103*, 269–281. [[CrossRef](#)]
11. Owusu, K.P.; Kuhn, D.C.S.; Bibeau, E.L. Capacitive Probe for Ice Detection and Accretion Rate Measurement: Proof of Concept. *Renew. Energy* **2013**, *50*, 196–205. [[CrossRef](#)]
12. Gao, H.; Rose, J.L. Ice Detection and Classification on an Aircraft Wing with Ultrasonic Shear Horizontal Guided Waves. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2009**, *56*, 334–344. [[CrossRef](#)] [[PubMed](#)]
13. Gómez Muñoz, C.Q.; García Márquez, F.P.; Sánchez Tomás, J.M. Ice Detection Using Thermal Infrared Radiometry on Wind Turbine Blades. *Measurement* **2016**, *93*, 157–163. [[CrossRef](#)]
14. Rizk, P.; Al Saleh, N.; Younes, R.; Ilinca, A.; Khoder, J. Hyperspectral Imaging Applied for the Detection of Wind Turbine Blade Damage and Icing. *Remote Sens. Appl. Soc. Environ.* **2020**, *18*, 100291. [[CrossRef](#)]
15. Gao, L.; Hong, J. Wind Turbine Performance in Natural Icing Environments: A Field Characterization. *Cold Reg. Sci. Technol.* **2021**, *181*, 103193. [[CrossRef](#)]
16. Shu, L.; Li, H.; Hu, Q.; Jiang, X.; Qiu, G.; McClure, G.; Yang, H. Study of Ice Accretion Feature and Power Characteristics of Wind Turbines at Natural Icing Environment. *Cold Reg. Sci. Technol.* **2018**, *147*, 45–54. [[CrossRef](#)]
17. Villalpando, F.; Reggio, M.; Ilinca, A. Prediction of Ice Accretion and Anti-Icing Heating Power on Wind Turbine Blades Using Standard Commercial Software. *Energy* **2016**, *114*, 1041–1052. [[CrossRef](#)]
18. Kreutz, M.; Alla, A.A.; Lütjen, M.; Ohlendorf, J.-H.; Freitag, M.; Thoben, K.-D.; Zimnol, F.; Greulich, A. Ice Prediction for Wind Turbine Rotor Blades with Time Series Data and a Deep Learning Approach. *Cold Reg. Sci. Technol.* **2023**, *206*, 103741. [[CrossRef](#)]
19. Xiao, J.; Li, C.; Liu, B.; Huang, J.; Xie, L. Prediction of Wind Turbine Blade Icing Fault Based on Selective Deep Ensemble Model. *Knowl.-Based Syst.* **2022**, *242*, 108290. [[CrossRef](#)]
20. Cheng, P.; Jing, H.; Hao, C.; Xinpan, Y.; Xiaojun, D. Icing Prediction of Fan Blade Based on a Hybrid Model. *Int. J. Perform. Eng.* **2019**, *15*, 2882. [[CrossRef](#)]
21. Ma, J.; Ma, L.; Tian, X. Wind Turbine Blade Icing Prediction Based on Deep Belief Network. In Proceedings of the 2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Hohhot, China, 24–26 October 2019; pp. 26–263.
22. Bai, X.; Tao, T.; Gao, L.; Tao, C.; Liu, Y. Wind Turbine Blade Icing Diagnosis Using RFECV-TSVM Pseudo-Sample Processing. *Renew. Energy* **2023**, *211*, 412–419. [[CrossRef](#)]
23. Tao, T.; Liu, Y.; Qiao, Y.; Gao, L.; Lu, J.; Zhang, C.; Wang, Y. Wind Turbine Blade Icing Diagnosis Using Hybrid Features and Stacked-XGBoost Algorithm. *Renew. Energy* **2021**, *180*, 1004–1013. [[CrossRef](#)]
24. Li, F.; Cui, H.; Su, H.; Ma, Z.; Zhu, Y.; Zhang, Y. Icing Condition Prediction of Wind Turbine Blade by Using Artificial Neural Network Based on Modal Frequency. *Cold Reg. Sci. Technol.* **2022**, *194*, 103467. [[CrossRef](#)]
25. Kreutz, M.; Alla, A.A.; Varasteh, K.; Ohlendorf, J.-H.; Lütjen, M.; Freitag, M.; Thoben, K.-D. Convolutional Neural Network with Dual Inputs for Time Series Ice Prediction on Rotor Blades of Wind Turbines. *Procedia CIRP* **2021**, *104*, 446–451. [[CrossRef](#)]
26. Albashish, D.; Hammouri, A.I.; Braik, M.; Atwan, J.; Sahran, S. Binary Biogeography-Based Optimization Based SVM-RFE for Feature Selection. *Appl. Soft Comput.* **2021**, *101*, 107026. [[CrossRef](#)]

27. Zeng, Z.; Cui, L.; Qian, M.; Zhang, Z.; Wei, K. A Survey on Sliding Window Sketch for Network Measurement. *Comput. Netw.* **2023**, *226*, 109696. [[CrossRef](#)]
28. Cai, J.; Wang, S.; Xu, C.; Guo, W. Unsupervised Deep Clustering via Contractive Feature Representation and Focal Loss. *Pattern Recognit.* **2022**, *123*, 108386. [[CrossRef](#)]
29. Chen, J.; Fu, C.; Xie, H.; Zheng, X.; Geng, R.; Sham, C.-W. Uncertainty Teacher with Dense Focal Loss for Semi-Supervised Medical Image Segmentation. *Comput. Biol. Med.* **2022**, *149*, 106034. [[CrossRef](#)]
30. Wu, Y.; Ma, X. A Hybrid LSTM-KLD Approach to Condition Monitoring of Operational Wind Turbines. *Renew. Energy* **2022**, *181*, 554–566. [[CrossRef](#)]
31. Cao, L.; Zhang, H.; Meng, Z.; Wang, X. A Parallel GRU with Dual-Stage Attention Mechanism Model Integrating Uncertainty Quantification for Probabilistic RUL Prediction of Wind Turbine Bearings. *Reliab. Eng. Syst. Saf.* **2023**, *235*, 109197. [[CrossRef](#)]
32. Abbaskhah, A.; Sedighi, H.; Akbarzadeh, P.; Salavatipour, A. Optimization of Horizontal Axis Wind Turbine Performance with the Dimpled Blades by Using CNN and MLP Models. *Ocean Eng.* **2023**, *276*, 114185. [[CrossRef](#)]
33. Jiang, G.; Yue, R.; He, Q.; Xie, P.; Li, X. Imbalanced Learning for Wind Turbine Blade Icing Detection via Spatio-Temporal Attention Model with a Self-Adaptive Weight Loss Function. *Expert Syst. Appl.* **2023**, *229*, 120428. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.