Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 4, 2667-2680
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i4.6638
© 2025 by the authors; licensee Learning Gate

Wind turbine blade icing prediction based on CNN-BiGRU with Optimized training strategy

Tao Chen[1], Xianghong Deng[2], Mingze Lei[1], Chonlatee Photong[1]*
[1]Faculty of Engineering, Mahasarakham University, MahaSarakham 44150, Thailand; chonlatee.p@msu.ac.th (C.P.).
[2]College of Electrical Engineering, Hunan Mechanical & Electrical Polytechnic, Changsha, 410151, China.

Abstract: Wind turbine blade icing significantly impacts the safety of wind farms and the efficiency of power generation, making precise and timely prediction a critical challenge. This study proposes an innovative deep learning framework that integrates Convolutional Neural Networks (CNNs) and Bidirectional Gated Recurrent Units (BiGRU) to enhance icing prediction. CNNs extract spatial features, while BiGRU captures temporal dependencies, enabling the model to effectively distinguish icing occurrences. To improve model optimization, cosine annealing was employed for dynamic learning rate adjustment, while cross-entropy loss was used to address class imbalance. Experimental results demonstrate that a 2-layer CNN architecture trained over 50 epochs achieves a balance between accuracy and computational efficiency, with CNN_2Layer-BiGRU attaining 96.55% accuracy and a 96.51% F1-score, outperforming traditional models. This approach reduces dependency on manual feature engineering, improves prediction accuracy and computational efficiency, and provides a foundation for an intelligent diagnostic system for wind turbine blade icing prediction.

Keywords: *Bidirectional Gated Recurrent Units, Convolutional Neural Networks, Cosine annealing, Cross-entropy loss, Wind turbine blade icing.*

1. Introduction

With the increasingly serious problems of carbon dioxide emissions, acid rain and energy shortage, wind energy, as a sustainable and environmentally friendly energy source that avoids the environmental pollution problems faced by traditional energy sources, has become a priority development direction for countries around the world. China is a vast country with vast grasslands and a long coastline and is rich in wind energy resources. However, these vast wind energy resources are primarily located in the cold northern regions and the humid, frigid southern areas, where environmental conditions are particularly harsh.

During wind turbine operation, especially at sub-zero temperatures, blade icing can occur when exposed to humid air, rain, salt spray or snow, and especially when exposed to cooling water droplets. As illustrated in Figure 1, captured using an Unmanned Aircraft System (UAS) imaging system at a mountainous wind farm in Hunan Province, China, blade icing can severely impact wind turbines. This is evident from the substantial increase in ice load on the blades due to icing accumulation. The uneven ice load on each blade increases the unbalanced load on the unit, which can affect the service life of the blades [1]. If the unit continues to operate, it may cause serious damage to the equipment; if the unit is shut down, it will significantly reduce the efficiency of wind power generation in low temperature areas. Blade icing changes the aerodynamic profile of the blades: due to the uneven thickness of the ice on all parts of the blades, the original aerodynamic properties of the blades are altered, which significantly reduces the load carrying capacity and power output of the wind turbine, thus significantly reducing the efficiency of the power generation [2]. In severe cases, the turbine may not even be able to start

History: Received: 10 February 2025; Revised: 15 April 2025; Accepted: 18 April 2025; Published: 28 April 2025
* Correspondence: chonlatee.p@msu.ac.th

generating electricity. When ice forms on the blade surface, it falls off as the temperature rises: this poses a safety threat to personnel and equipment on site and increases operational risks. Therefore, developing wind turbine blade icing prediction models is crucial for practical applications in industry, helping companies enhance operational efficiency and minimize risks.

Figure 1.
Icing Condition of Hunan Wind Farm, China.

Currently, wind turbine blade icing diagnosis methods are generally classified into direct and indirect approaches [3] . The direct method directly monitors the blade surface through sensors or equipment, and commonly used technologies include blade icing prediction system [4] infrared optical image processing [5] and hyperspectral imaging [6] etc. These methods have high real-time and accuracy and can directly reflect the icing status of the blade. These methods have high real-time performance and accuracy and can directly reflect the icing status of the blade. However, direct methods usually require additional hardware equipment, which is costly and less reliable in harsh environments, making it difficult to be widely applied to large-scale wind farms.

In contrast, the indirect method diagnoses by analyzing the relationship between blade characteristics and icing features, which is further classified into physical model-based and data-driven approaches. Tao, et al. [7] extracted hybrid features reflecting both short-term and long-term icing effects based on potential physical icing accumulation processes and developed a stacked-XG Boost model for blade icing diagnosis. Jiménez, et al. [8] introduced a pattern recognition approach utilizing guided ultrasound and machine learning to detect icing on wind turbine blades. While Xiang, et al. [9] proposed a new fault prediction and diagnosis system by combining ultrasound technology with wavelet transform. Physical model-based methods rely on small-scale experiments and mathematical modeling, and the results are difficult to generalize to large-scale wind turbine applications.

The data-driven approach leverages operational data from the Supervisory Control and Data Acquisition (SCADA) system and integrates machine learning or deep learning techniques to extract icing features, making it especially suitable for large-scale wind turbine applications Ye and Ezzat [10]. Muñoz, et al. [11] proposed a multiscale wavelet-driven transformer (MWT)-based method, which integrates Discrete Wavelet Decomposition (DWD) and Transformer Blocks, to extract multiscale features from wind turbine sensor data from wind turbine sensor data to extract multi-scale features. Additionally, Cheng, et al. [12] introduced a deep category imbalance semi-supervised (DCISS) model to estimate wind turbine blade icing. The model combines category imbalance learning with semi-supervised learning (SSL) and enhances feature extraction with a channel-calibrated attention module.

In data-driven approaches, machine learning models such as deep fully connected neural networks (FCNN) [13] support vector machines and random forest models [14] and models based on convolutional neural networks (CNNs) and bidirectional long- and short-term memory networks (BiLSTMs) [15] have been applied to leaf icing prediction with some success. With the deepening of research, more complex deep learning models have been proposed, which further advance the accuracy of diagnostic methods. Lai, et al. [16] proposed an efficient and accurate leaf icing diagnostic method using a hybrid feature extraction method including recursive feature elimination (RFE) and sliding window algorithms, which was validated in combination with the B-SMOTE-Bi-GRU model. Kreutz, et al. [17] proposed a selective deep integration method based on the Selective Deep Integration (GSDE) model of GMDH, which combines multiple neural networks to form multiple deep neural networks based on focal loss for wind turbine blade icing prediction. In addition, Li, et al. [18] proposed a method combining the Feature Selection algorithm and the 1D-CNN-SBiGRU model to solve the high-dimensional and unbalanced data problems through feature selection and data reconstruction.

Although existing methods are excellent in improving accuracy, many complex features engineering and multi-model integration methods require long computation time, which is difficult to meet the needs of wind farms for efficiency and fast response. Therefore, this paper introduces a deep learning model that integrates a convolutional neural network (CNN) with a bi-directional gated recurrent unit (BiGRU) to enhance short-term wind turbine blade icing prediction. The model accurately models the icing state by efficiently extracting the spatial features through CNN and capturing the time series relationship through BiGRU. By incorporating the cosine annealing learning rate and cross-entropy loss function, the training process is optimized, leading to significant improvements in accuracy and efficiency. This approach ensures real-time prediction capabilities and meets enterprise demands for fast and effective computation.

2. Relevant Studies

2.1. CNN Algorithm

As a variant of feedforward neural networks, Convolutional Neural Networks (CNNs) have been increasingly applied in natural language processing (NLP) research in recent years [19]. A CNN primarily consists of three key components: the input layer, convolutional and pooling layers, and the fully connected layer [20].

Feature extraction in the convolutional layer begins by encoding text as a word vector matrix, which is then processed using convolutional filters of various kernel sizes. These filters operate with fixed parameter values during the scanning process, ensuring consistency in feature prediction. The output is a feature map, where each element is generated through filtering with identical convolutional parameters. The structure of the CNN is illustrated in Figure 2.

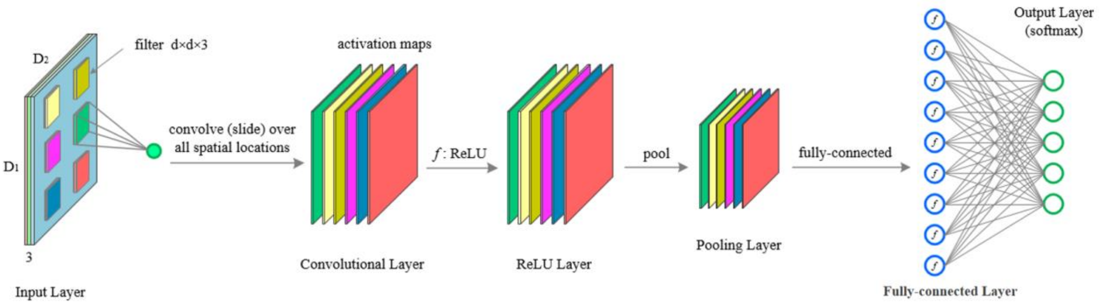

Figure 2.
CNN Network Architecture [21].

In convolutional neural networks (CNNs), the increase in the number of layers has a significant effect on the feature extraction ability and overall performance of the model [22]. Generally, increasing the number of layers can improve the network's ability to extract higher-order features, but too many layers may lead to problems such as overfitting or gradient vanishing [23] . Therefore, choosing the appropriate number of layers is crucial for model performance.

For multilayer CNN, the layer-by-layer process of feature extraction is expressed as follows.

$$H^{(L)} = f(W^{(L)} * f(W^{(L-1)} * \cdots f(W^{(1)} * H^0 + b^1) + \cdots b^{(L-1)}) + b^{(L)}) \qquad (1)$$

Where L represents the total number of convolutional layers, $\mathrm{H}^{(\mathrm{L})}$ denotes the output feature map of the L^{th} layer, $\mathrm{W}^{(\mathrm{L})}$ is the convolutional kernel weight matrix, and $\mathrm{b}^{(\mathrm{L})}$represents the bias term. The function f() applies a nonlinear activation, while *signifies the convolution operation.

2.2. The BIGRU Algorithm

The Gated Recurrent Unit (GRU), a variant of the Recurrent Neural Network (RNN), is specifically designed to address the issues of gradient vanishing and explosion that occur in standard RNNs when processing long sequential data [24]. GRUs regulate information flow through gating mechanisms, such as update and reset gates, enabling more efficient capture of long-term dependencies while minimizing computational complexity [25]. However, the unidirectional structure of the traditional GRU can only learn from time-series historical information [26] and cannot utilize future information, while in wind turbine icing prediction, the time-series nature of SCADA data needs to be fully explored for contextual dependencies [21].

For this reason, Bidirectional GRU (BiGRU for short) is introduced. BiGRU enhances the model's ability to capture temporal dependencies by incorporating both forward and backward GRU sub-networks, allowing it to utilize past and future information for a more comprehensive representation of time series features. The core of BiGRU consists of update gates and reset gates, which are used for the control of the state updating and forgetting information, and its bi-directional structure is able to synthesize the feature representations of the front and back time steps.

For the input time series X={ $x_1, x_2, \ldots, x_{T-1}, x_T$}, BiGRU is computed as follows.

$$h_t = [\vec{h}_t; \overleftarrow{h}_t], t = 1,2 \cdots T \qquad (2)$$

Where GRU(x_t,$\mathrm{h}_{\mathrm{t-1}}$) represents the hidden state of the forward GRU, andGRU(x_t,$\mathrm{h}_{\mathrm{t+1}}$) denotes the hidden state of the reverse GRU. The final output is obtained by concatenating these forward and backward hidden states. The hidden state sequence H = {h_1,h_2,...,h_T}, extracted by BiGRU, effectively captures bi-directional contextual information from the time series. This comprehensive representation is then mapped to the predicted icing state through a fully connected layer, enabling precise wind turbine icing prediction.

2.3. Joint CNN-BiGRU Mechanism

This paper presents a wind turbine icing prediction model that integrates CNN with Bi-directional GRU to enhance predictive performance.CNN is used to extract spatial features of wind turbine SCADA data, while BiGRU processes time series data through its bi-directional structure to capture forward and backward contextual information for better understanding of the temporal properties of the data. CNN firstly performs the input data Convolutional operation is performed to extract the low-level features and reduce the dimensionality through pooling layer to retain the important spatial information. Then, BiGRU partially extracts more comprehensive features from the past and future information of the time series to improve the model's ability to capture temporal dependencies.

The architecture of the model (e.g., Figure 3) consists of five main components: an input layer, a convolutional layer, a BiGRU layer, an output layer, and a final predicted output. The data is first passed through a CNN layer for feature extraction, followed by a BiGRU layer to capture the timing information, and finally an output layer to generate the prediction results.The model's output is typically a fully connected layer that transforms extracted features into target predictions.

$$\hat{y} = W_{OUT} \cdot h + b_{out} \qquad (3)$$

Where $\hat{y}$ represents the predicted output, W_{OUT} is the weight matrix of the output layer, h denotes the features extracted from previous layers (CNN and BiGRU), and b_{out} is the bias term of the output layer.

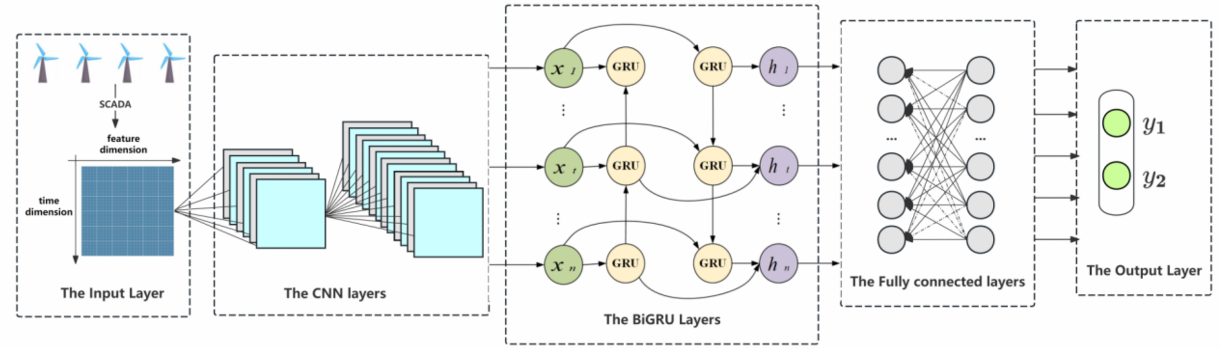

Figure 3.
CNN-BiGRU Algorithm Architecture [27].

The joint CNN-BiGRU mechanism demonstrates the flow of data from input to feature extraction to final output, fully utilizing the advantages of convolutional networks in spatial feature extraction and BiGRU in time series modeling.

2.4. Optimization Model and Loss Function Selection

During debugging, the cosine annealing learning rate scheduler is introduced to dynamically adjust the learning rate, enhancing training efficiency and model performance. Cosine annealing, inspired by Simulated Annealing, gradually decreases the learning rate, helping the model converge more smoothly toward the global optimal solution in the final stages of training [28].
The cosine annealing learning rate equation is given by.

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + \cos\left(\frac{t}{T_{max}}\pi\right)\right) \qquad (4)$$

The learning rate at the current training step is denoted as η_t, with η_{max} and η_{min} representing its upper and lower bounds, respectively. Here,t indicates the current training step, while T_{max} defines the annealing period, specifying the total number of steps over which the learning rate is adjusted.

In the icing prediction task, in order to measure the difference between the model's predicted probability distribution and the actual categories, we introduce the Cross Entropy Loss function as the optimization objective function [29]. Since the icing problem can be regarded as a binary categorization problem of icing and non-icing, the properties of the Cross Entropy Loss function are well suited to the needs of such a task. Specifically, the Cross Entropy Loss function can accurately calculate the difference between the model's predicted probability distribution and the true category labels to help the model learn the classification boundaries better, especially in the face of category imbalance, and can effectively avoid the model's prediction biased towards the large category (i.e., the non-icing state).

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \log(\hat{y}_{ij}) \qquad (5)$$

Here, N represents the total number of samples, while C denotes the number of categories within the icing classification. In the wind turbine icing task, this classification is simplified into a binary distinction between icing and non-icing, setting C=2. The variable y_{ij} indicates the true label of sample i for category j, and $\hat{y}_{ij}$ represents the model's predicted probability for category j.

Moreover, the cross-entropy loss function helps minimize the misclassification of icing state samples by refining prediction probabilities. In wind turbine blade icing prediction, it enhances the learning of rare icing cases by adjusting the weights of smaller errors, ultimately improving the model's accuracy and robustness in real-world applications. Integrating the cosine annealing learning rate scheduler, the cross-entropy loss function enhances the training process, ensuring efficient learning at each stage and leading to more stable and precise predictions.

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

Serial Number	Variable Name	Description	Serial Number	Variable Name	Description
1	Wind_speed	Wind speed (m/s)	9	Pitch_angle (1,2,3)	Blade pitch angle (°)
2	Generator_speed	Generator speed (rpm)	10	Pitch_moto_tmp (1,2,3)	Pitch motor temperature (°C)
3	Power (kw)	Generated power (kW)	11	Acc_xx	Nacelle X-axis acceleration (m/s^2)
4	Wind_direction (°)	Wind direction (°)	12	Acc_yy	Nacelle Y-axis acceleration (m/s^2)
5	Wind_direction_ mean	Average wind direction (°)	13	Environment_tmp	Ambient temperature (°C)
6	Yaw_position	Yaw position (°)	14	Int_tmp	Internal nacelle temperature (°C)
7	Yaw_speed	Yaw speed (°/s)	15	Pitch_ng5_tmp (1,2,3)	NG5 pitch system temperature (°C)
8	Pitch_speed (1,2,3)	Blade pitch speed (°/s)	16	Pitch_ng5_DC (1,2,3)	NG5 pitch DC bus voltage (V)

3. Arithmetic Simulation and Analysis

3.1. Input Feature Selection

This study gathered wind turbine blade icing data from a wind farm located in Hunan Province, China, utilizing turbines produced by Harbin Electric Group. The SCADA data was recorded from December 1 to December 30, with a 7-second data sampling interval, covering multiple monitoring variables. From hundreds of sensor data, the engineers were able to identify 26 key variables related to blade icing and pinpointed when the icing occurred. Specific variable information is detailed in Table 1.

3.2. Data Pre-Processing

To enhance the accuracy of wind turbine blade icing forecasting and minimize the impact of redundant features on the model, this study introduces a feature-enhanced deep learning approach for predictive analysis. First, comprehensive preprocessing is performed on the SCADA data, including operations such as outlier removal, normalization, normalization, and data labeling and segmentation. Subsequently, a convolutional neural network (CNN) is used to extract power features and transform them into power mean square error features to achieve feature enhancement. Next, the CNN-BiGRU deep learning model is developed by integrating the spatial feature extraction capability of CNN with the temporal modeling strength of BiGRU. Through analyzing and diagnosing the spatio-temporal characteristics of wind turbine blade operation data, the model effectively predicts blade icing conditions with high accuracy. The detailed process is illustrated in Figure 4.

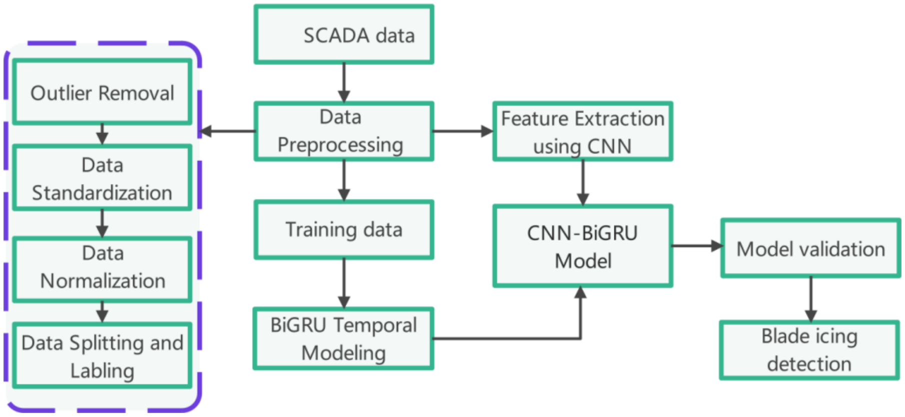

Figure 4.
Blade Ice Prediction Flowchart [13].

The data utilized in this study was obtained from the SCADA system; however, it contained anomalies, redundant entries, and missing values. In order to mitigate the impact of these issues on model training and evaluation, the following integrated data preprocessing process was implemented:

(1) Data reading and preliminary screening

Iterate through all files in the data directory and perform preliminary screening based on specific keywords in the file names. Only eligible CSV files are retained for further analysis.

(2) Data Labeling and Missing Value Filling

Experienced wind turbine engineers labeled the raw data as either "normal" or "iced". To ensure data continuity and integrity, missing values were filled in using the *fillna()* method. This prevents gaps in the time series from affecting subsequent analysis and model training.

(3) Data Standardization and Segmentation

Since SCADA data are time series data collected from sensors at fixed intervals, they need to be standardized and segmented to meet the input requirements of the CNN-BiGRU wind turbine blade icing prediction model:

- Standardization: Min-Max standardization method is used to scale the feature values to the range of [0, 1] to reduce the impact of unit and magnitude differences between different features on the model. The standardization formula is as follows:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \qquad (6)$$

Where X' represents the pre-processed feature data, X refers to the original data, and X_{max} and X_{min} indicate the maximum and minimum feature values, respectively.

- Segmentation: Segment the normalized time series data into fixed length segments (500 data points per segment) to fit the input format of the model.

(4) Icing data processing

The same preprocessing steps, including segmentation and normalization, were applied to the icing data. In addition, fault data from other data sources are extracted, preprocessed using a consistent methodology, and integrated into the overall dataset.

Through the preprocessing steps described above, the SCADA data is converted into a structured, clean, and standardized format suitable for model training and prediction, thereby improving the reliability and accuracy of the wind turbine blade icing prediction system.

3.2. Evaluation Metrics

This study evaluates the proposed model using unbalanced data, where the number of abnormal (icing) samples is much smaller than the number of normal (non-icing) samples. This imbalance may lead to inflated accuracy scores, thus making the model evaluation inaccurate.Since this study

DOI: 10.55214/25768484.v9i4.6638

concentrates on wind turbine blade icing, where icing and non-icing conditions are treated as positive and negative samples, respectively, Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC) are adopted as performance evaluation metrics. Their definitions are as follows:

$$Pr\ e\ cision = \frac{TP}{TP+FP} \tag{7}$$

$$Recall=\frac{TP}{TP+FN} \tag{8}$$

$$F1 = \frac{2\times Pr\ ecision\times Re\ call}{Pr\ ecision+Re\ call} \tag{9}$$

$$ACC=\frac{TP+TN}{TP+FP+TN+FN} \tag{10}$$

$$MCC = \frac{TP\times TN-FP\times FN}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \tag{11}$$

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

4. Experiments and Discussions

4.1. Baseline Comparison

In order to verify the validity of the proposed model, we selected several common time series modeling methods as baseline models for comparison, including LSTM, BiLSTM, GRU, BiGRU and CNN-BiGRU. The reasons for selecting these baseline models are as follows:

LSTM (Long Short-Term Memory) [30]: LSTM, a well-established Recurrent Neural Network (RNN) variant, is specifically designed to address long-term dependency challenges in sequential data. In this study, a 2-layer LSTM architecture is utilized, with hidden units selected from (8, 16, 32, 64). The model with the highest performance is ultimately chosen for evaluation.

BiLSTM (Bidirectional LSTM) [31]: BiLSTM extends LSTM by incorporating a bidirectional mechanism, enabling the model to learn both past and future contextual information simultaneously. The same experimental setup as LSTM is applied for comparison.

GRU (Gated Recurrent Unit) and BiGRU (Bidirectional GRU): These models, described in Section 2.2, adopt gating mechanisms to optimize memory usage and improve computational efficiency in sequential learning tasks.

CNN-BiGRU (Convolutional and Bidirectional GRU Combined Model): This model integrates CNN for spatial feature extraction and BiGRU for temporal sequence modeling, enhancing the representation of spatio-temporal dependencies. By effectively capturing complex patterns in large datasets and short time series, CNN-BiGRU demonstrates strong applicability in wind turbine blade icing prediction.

The model training and evaluation were conducted on a Windows 10 platform utilizing the PyTorch 1.8.1 deep learning framework, with Python 3.8 as the programming environment. The corresponding experimental results are presented in Table 2.

Table 2.
Baseline model comparison results.

Model	Accuracy	Precision	Recall	F1-score	MCC
CNN	77.6%	77.6%	77.8%	77.6%	55.4%
LSTM	76.4%	77.7%	75.4%	75.6%	53.1%
BiLSTM	74.1%	74.1%	74.3%	74.1%	48.4%
GRU	69.5%	73.2%	71.0%	69.1%	44.2%
BiGRU	81.6%	81.9%	81.1%	81.3%	63.0%
CNN-BiGRU	85.6%	85.6%	85.9%	85.6%	71.5%

From the results, it can be seen that BiGRU has the best overall performance with an Accuracy of 85.6%, which is better than other time series models. This indicates that BiGRU can capture the time-series features of SCADA data more effectively, which is more advantageous for the task of wind turbine blade icing prediction. However, the performance of the traditional RNN structure on SCADA data still has some limitations in terms of insufficient feature extraction capability and high model complexity. Therefore, we further explore the performance of different deep learning architectures in feature extraction (see 4.3).

4.2. Sensitivity Experiments

To further assess the influence of various factors on the effectiveness of the wind turbine blade icing prediction model, two sensitivity analyses were performed: examining the effect of a shallow CNN architecture on model performance and evaluating the impact of training iterations on overall accuracy.

4.2.1. Impact Of Shallow CNN Architecture on Model Performance

In order to investigate the performance of CNN networks with different depths in wind turbine blade icing prediction, we designed a CNN model containing 1, 2 and 3 convolutional layers and conducted sensitivity experiments. The experimental results are shown in Table 4, with the increase of the number of CNN layers, there is a significant improvement in the F1-score, but the balance between the performance and the computational cost is of concern.

Table 3.
Comparison of Performance and Computational Cost for Different CNN Layer Depths.

Number of Layers	Accuracy	F1-score	Number of Parameters	Training Time (seconds/epoch)
CNN_1Layer	94.83%	94.79%	4354	5
CNN_2Layer	96.55%	96.51%	16706	8
CNN_3Layer	95.98%	95.92%	29058	23

1-layer CNN. This model has the lowest computational cost (training time 5 sec/epoch), but the F1-score is only 94.79%, which indicates that it is deficient in the balance of precision and recall and is prone to misprediction or omission.

2-layer CNN. Performs best among all models with an F1-score of 96.51%, while the accuracy is improved to 96.55%. The model strikes an optimal balance between accuracy and recall, and the training time is relatively controllable (8 sec/epoch) for the wind turbine icing prediction task.

3-layer CNN. After further increasing the number of CNN layers, the F1-score is improved (95.92%), but the increase is limited, while the computational overhead is significantly increased (training time 23 sec/epoch). Although the performance of the model is still high, its advantage over 2-layer CNN is not obvious considering the training cost.

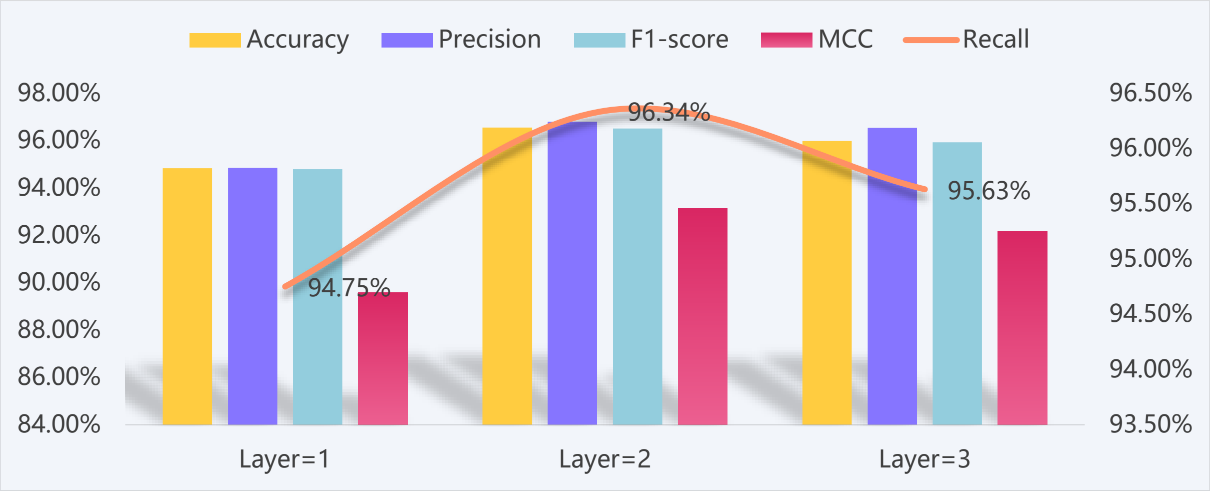

Figure 5.
Comparison of Performance and Computational Cost for Different CNN Layer Depths.

DOI: 10.55214/25768484.v9i4.6638

From the visualization results in Fig. 5, it can be seen that 2-layer CNN achieves the best performance in terms of accuracy, F1-score, precision and recall, with the F1-score as high as 96.51%. In contrast, the accuracy and recall of 1-layer CNN are low, which is difficult to meet the actual demand, while 3-layer CNN has some improvement, but the computational cost increases more, which is less cost-effective. Therefore, in the wind turbine blade icing prediction task, 2-layer CNN achieves the best balance between performance and computational cost and is a more ideal choice.

4.2.2. Effect of Training Rounds on Model Performance

To further explore the impact of training iterations (Epochs) on model performance, experiments were conducted with Epoch values of 30, 50, and 80, assessing Accuracy, Precision, Recall, F1-score, and training duration. The corresponding experimental results are presented in the following table.

Table 4.
Impact of Epochs on Model Performance and Training Time.

Epoch	Accuracy	Precision	Recall	F1-score	Training Time (minutess)
30	62.06%	70.99%	64.43%	59.72%	7
50	94.25%	95.19%	93.75%	94.14%	12
80	96.55%	96.79%	96.34%	96.51%	31

In addition, Figure 6 depicts the evolution of the loss function over varying epochs, offering a comprehensive assessment of the model's convergence properties and generalization performance. Experimental results indicate that an increase in training iterations significantly impacts both model accuracy and computational complexity.

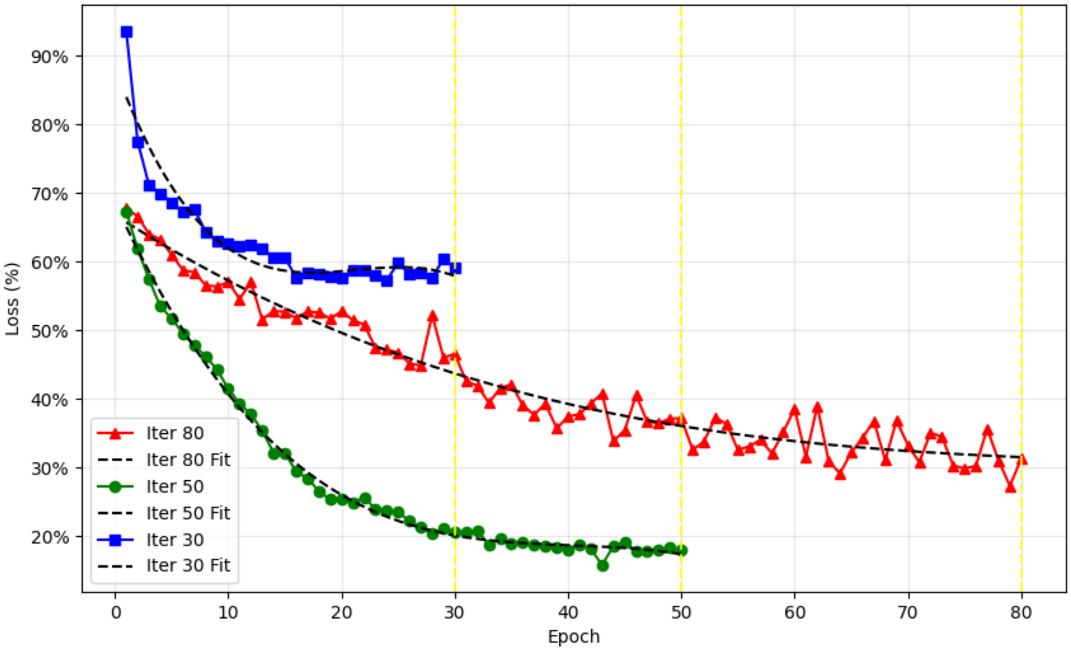

Figure 6.
Loss Curve for Different Epochs during Model Training.

(1) Epoch=30

The training time is only 7 minutes, but the accuracy (62.06%) and F1-score (59.72%) of the model are low, indicating insufficient training, limited feature extraction capability, and unsatisfactory model performance on the classification task.

(2) Epoch=50

The accuracy (94.25%) and recall (93.75%) of the model are greatly improved, and the F1-score reaches 0.9414, which indicates that the model has basically converged and is able to distinguish the samples efficiently, meanwhile the training time is controlled to be 12 minutes, which is within the permissible range of the computational resources, and it has a high practicability.

(3) Epoch=80

The model performance is further improved (accuracy 96.55%, F1-score 96.51%), but the gain compared to Epoch=50 is limited. In addition, the training time increased significantly to 31 minutes, with a substantial increase in computational cost. From the loss curves, there were more obvious fluctuations during the training process, indicating that the model may have entered the overfitting stage and the generalization ability decreased.

Taking into account the model's performance, computational cost, and generalization capability,Epoch=50 achieves an optimal balance between performance and computational resource consumption. On the one hand, it can effectively reduce the risk of overfitting, while ensuring the stability and reliability of the model in the task of wind turbine blade icing prediction. Therefore, in the subsequent experiments, we choose Epoch=50 as the optimal number of training rounds to ensure the model's generalization ability and practical application value.

4.3. Deep Learning Architecture Comparison

IIn Section 4.2.1, we examine how the number of CNN layers impacts the effectiveness of wind turbine blade icing prediction. The experimental findings indicate that increasing the CNN layers (to 2 or 3) enhances the model's performance. However, as a feature extraction tool, CNN is limited by the design of network structure, and there is still room for optimization. To further explore better feature extraction strategies, we introduce multiple deep learning architectures to evaluate the suitability of diverse models for this task.

In this experiment, six distinct architectures were selected to comprehensively explore the temporal information and complex relationships within SCADA data. These architectures encompass a variety of feature extraction methods, including temporal attention, graph neural networks, and residual learning. The specific experimental results are presented in Table 5.

TAPNet (Temporal Attention Network) [32]: This model employs a temporal attention mechanism to enhance focus on critical time points, improving prediction accuracy by assigning adaptive weights to input data from different time steps.

GTAN (Graph Temporal Attention Network) [33]: By integrating Graph Neural Networks (GNN) with Temporal Attention Mechanisms, GTAN effectively captures correlations between turbines, thereby improving the prediction of icing risks in multi-turbine systems.

CBAM (Convolutional Block Attention Module) [34] : CBAM incorporates channel and spatial attention mechanisms into the CNN architecture, enabling the model to prioritize critical features while suppressing irrelevant noise. This design demonstrates exceptional performance in icing state prediction.

ResNet (Residual Network) [35]: ResNet addresses the gradient vanishing problem in deep networks through residual connections, allowing the model to efficiently learn complex features from SCADA data and better model the dynamics of wind turbine blade icing processes.

MLP (Multi-Layer Perceptron) [36] : Although MLP, composed of fully connected layers, exhibits limited capability in time series modeling, it remains effective in capturing underlying data trends and serves as a reliable benchmark model.

CNN_2Layer-BiGRU (Convolutional and Bidirectional GRU Combined Model): This model combines CNN for spatial feature extraction with BiGRU for temporal information processing. The integration enhances the model's ability to represent both spatial and temporal features, making it highly suitable for wind turbine blade icing prediction.

Table 5.
Comparison of Performance Across Different Deep Learning Architectures.

Model	Accuracy	Precision	Recall	F1-score	MCC
TAPNet	67.82%	69.08%	68.63%	67.75%	69.19%
GTAN	84.48%	84.48%	84.71%	84.46%	37.70%
CBAM	69.64%	66.75%	65.97%	64.75%	32.71%
ResNet	91.38%	91.36%	91.28%	91.31%	82.64%
MLP	85.63%	85.74%	85.96%	85.62%	71.70%
CNN_2Layer-BiGRU	96.55%	96.79%	96.34%	96.51%	93.13%

The experimental results discovered indicate that the CNN_2Layer-BiGRU model achieves the best index performance among the four supervised models. The accuracy is 96.55% with recall and precision of above 96%, which is clearly the best chance among others. Hence, it is obvious the approach of joint CNN and BiGRU exhibits remarkable benefits in SCADA data time-series modeling and feature selection, and it also provides a more efficient algorithm for detecting blade icing in wind turbines.

Moreover, ResNet is one of the best-trained approaches, with an accuracy of 91.38%. This further proves that the deep CNN features extraction is the main strength of this structure. GTAN and TAPNet show a comparative gain in Recall due to the fact that they are tailored for temporal attention, but still fail to reach the performance level of CNN-BiGRU.

5. Conclusions

Icing of wind turbine blades affects the stable operation and power generation efficiency of wind farms, and the establishment of an efficient prediction model is crucial to ensure the safety of wind turbines and improve power generation efficiency. Considering the actual demands of enterprises, we deal with the deep learning architecture performance on the confusing task and are also concerned with the comparison of CNN structure with training rounds, as well as the optimization strategy on model performance. Results indicate that the 2-layer CNN structure yields the best results on feature extraction and computation, while the number of training rounds to be set to 50 yields the best outcomes in terms of both accuracy and generalization capability. Under this situation, the model does not suffer the problem of either under-training or over-fitting.

To improve the model performance, this study applies cosine annealing learning rates to schedule the training process and cross-entropy loss function for the category imbalance data rebalancing. The performance of the models on the experimental data shows that CNN_2Layer-BiGRU outperforms traditional RNN, CNN, and Attention Mechanism models with a best recording of 96.55% and 96.51% for model accuracy and F1-score levels. These results indicate that CNN and BiGRU are capable enough in spatial feature extraction and to capture time series variations for icing prediction results enhancement.

All in all, the CNN-BiGRU model integrated with the refined parameter tuning strategy produced the best results in ice prediction on SCADA data, which is not only considering the model efficiency but also the predictive ability and the algorithm stability.

The results provide an efficient and stable intelligent diagnosis scheme for wind turbine blade icing prediction, which helps to improve the safety and operational efficiency of wind farms.

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.

Copyright:

References

[1] R. Wang, H. Qiu, G. Jiang, X. Liu, and X. Cheng, "Class-Imbalanced Spatial–Temporal Feature Learning for Blade Icing Recognition of Wind Turbine," *IEEE Transactions on Industrial Informatics*, 2024. https://doi.org/10.1109/TII.2024.3393550

[2] 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

[3] 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

[4] 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

[5] Z. Zhang and Z. Shu, "Unmanned aerial vehicle (UAV)-assisted damage detection of wind turbine blades: A review," *Energies*, vol. 17, no. 15, p. 3731, 2024. https://doi.org/10.3390/en17153731

[6] P. Rizk, F. Rizk, S. S. Karganroudi, A. Ilinca, R. Younes, and J. Khoder, "Advanced wind turbine blade inspection with hyperspectral imaging and 3D convolutional neural networks for damage detection," *Energy and AI*, vol. 16, p. 100366, 2024. https://doi.org/10.1016/j.egyai.2024.100366

[7] 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

[8] A. A. Jiménez, F. P. G. Márquez, V. B. Moraleda, and C. Q. G. Muñoz, "Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis," *Renewable energy*, vol. 132, pp. 1034-1048, 2019. https://doi.org/10.1016/j.renene.2018.08.050

[9] D. Xiang, X. Cheng, S. Chen, M. Liu, F. Shi, and X. Liu, "Multiscale Wavelet-Driven Transformerfor Blade Icing Detection," in *Proceedings of the 2024 3rd International Conference on Frontiers of Artificial Intelligence and Machine Learning*, 2024, pp. 204-210, doi: https://doi.org/10.1145/3653644.3653650

[10] F. Ye and A. A. Ezzat, "Icing detection and prediction for wind turbines using multivariate sensor data and machine learning," *Renewable Energy*, vol. 231, p. 120879, 2024. https://doi.org/10.1016/j.renene.2024.120879

[11] C. Q. G. Muñoz, A. A. Jiménez, and F. P. G. Márquez, "Wavelet transforms and pattern recognition on ultrasonic guides waves for frozen surface state diagnosis," *Renewable Energy*, vol. 116, pp. 42-54, 2018. https://doi.org/10.1016/j.renene.2017.03.052

[12] 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

[13] 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

[14] S. Mao, Y. Wu, L. Zhao, and W. Mo, "Surrogate modelling for ice accretion prediction on wind turbine blades based on support vector regression," *IET Renewable Power Generation*, vol. 18, no. 4, pp. 567–575, 2024. https://doi.org/10.1049/rpg2.12999

[15] J. Yan, Y. Liu, L. Li, and X. Ren, "Wind Turbine Condition Monitoring Using the SSA-Optimized Self-Attention BiLSTM Network and Changepoint Detection Algorithm," *Sensors*, vol. 23, no. 13, p. 5873, 2023. https://doi.org/10.3390/s23135873

[16] Z. Lai, X. Cheng, X. Liu, L. Huang, and Y. Liu, "Multiscale wavelet-driven graph convolutional network for blade icing detection of wind turbines," *IEEE Sensors Journal*, vol. 22, no. 22, pp. 21974-21985, 2022. https://doi.org/10.1109/jsen.2022.3211079

[17] M. Kreutz *et al.*, "Ice prediction for wind turbine rotor blades with time series data and a deep learning approach," *Cold Regions Science and Technology*, vol. 206, p. 103741, 2023. https://doi.org/10.1016/j.coldregions.2022.103741

[18] Y. Li *et al.*, "Prediction of wind turbine blades icing based on feature Selection and 1D-CNN-SBiGRU," *Multimedia Tools and Applications*, vol. 81, no. 3, pp. 4365-4385, 2022. https://doi.org/10.1007/s11042-021-11700-7

[19] M. Ekman, *Learning deep learning: Theory and practice of neural networks, computer vision, natural language processing, and transformers using TensorFlow*. Addison-Wesley Professional, 2021.

[20] A. Zafar *et al.*, "A comparison of pooling methods for convolutional neural networks," *Applied Sciences*, vol. 12, no. 17, p. 8643, 2022. https://doi.org/10.3390/app12178643

[21] J. Kang, X. Zhu, L. Shen, and M. Li, "Fault diagnosis of a wave energy converter gearbox based on an Adam optimized CNN-LSTM algorithm," *Renewable Energy*, vol. 231, p. 121022, 2024. https://doi.org/10.1016/j.renene.2024.121022

[22] N. Singh, V. Tewari, and P. Biswas, "Vision transformers for cotton boll segmentation: Hyperparameters optimization and comparison with convolutional neural networks," *Industrial Crops and Products*, vol. 223, p. 120241, 2025.

[23] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS journal of photogrammetry and remote sensing*, vol. 173, pp. 24-49, 2021. https://doi.org/10.1016/j.isprsjprs.2020.12.010

[24] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," *Information*, vol. 15, no. 9, p. 517, 2024. https://doi.org/10.20944/preprints202408.0748.v1

[25] W. Zheng and G. Chen, "An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13902-13914, 2021. https://doi.org/10.1109/tcyb.2021.3121312

[26] W. Tian, X. Cheng, F. Shi, G. Li, S. Chen, and H. Zhang, "Gated Convolutional Neural Network for Wind Turbine Blade Icing Detection," in *2022 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, 2022: IEEE, pp. 186-191, doi: https://doi.org/10.1109/rcar54675.2022.9872273.

[27] 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

[28] T. Guilmeau, E. Chouzenoux, and V. Elvira, "Simulated annealing: A review and a new scheme," in *2021 IEEE statistical signal processing workshop (SSP)*, 2021: IEEE, pp. 101-105, doi: https://doi.org/10.1109/ssp49050.2021.9513782

[29] A. Mao, M. Mohri, and Y. Zhong, "Cross-entropy loss functions: Theoretical analysis and applications," in *International conference on Machine learning*, 2023: PMLR, pp. 23803-23828.

[30] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020. https://doi.org/10.1016/j.physd.2019.132306

[31] R. L. Abduljabbar, H. Dia, and P.-W. Tsai, "Unidirectional and bidirectional LSTM models for short-term traffic prediction," *Journal of Advanced Transportation*, vol. 2021, no. 1, p. 5589075, 2021. https://doi.org/10.1155/2021/5589075

[32] X. Zhang, Y. Gao, J. Lin, and C.-T. Lu, "Tapnet: Multivariate time series classification with attentional prototypical network," in *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 04, pp. 6845-6852, doi: https://doi.org/10.1609/aaai.v34i04.6165

[33] L. Ying, Z. Xu, H. Zhang, J. Xu, and X. Cheng, "Graph Temporal Attention Network for Imbalanced Wind Turbine Blade Icing Prediction," *IEEE Sensors Journal*, 2024. https://doi.org/10.1109/jsen.2024.3358873

[34] Q. Lu, W. Ye, and L. Yin, "ResDenIncepNet-CBAM with principal component analysis for wind turbine blade cracking fault prediction with only short time scale SCADA data," *Measurement*, vol. 212, p. 112696, 2023. https://doi.org/10.1016/j.measurement.2023.112696

[35] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International joint conference on neural networks (IJCNN)*, 2017: IEEE, pp. 1578-1585, doi: https://doi.org/10.1109/ijcnn.2017.7966039

[36] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN)," *Clinical eHealth*, vol. 4, pp. 1-11, 2021. https://doi.org/10.1016/j.ceh.2020.11.002