



ARTICLE

Gated Fusion Based Transformer Model for Crack Detection on Wind Turbine Blade

Wenyang Tang^{1,2}, Cong Liu^{1,*} and Bo Zhang²

¹School of Mechanical Engineering, Yancheng Institute of Technology, Yancheng, 224007, China

²Huaneng Offshore Wind Power Science and Technology Research Co., Ltd., Yancheng, 224007, China

*Corresponding Author: Cong Liu. Email: liucong12601@126.com

Received: 29 March 2023 Accepted: 30 May 2023 Published: 31 October 2023

ABSTRACT

Harsh working environments and wear between blades and other unit components can easily lead to cracks and damage on wind turbine blades. The cracks on the blades can endanger the shafting of the generator set, the tower and other components, and even cause the tower to collapse. To achieve high-precision wind blade crack detection, this paper proposes a crack fault-detection strategy that integrates Gated Residual Network (GRN), a fusion module and Transformer. Firstly, GRN can reduce unnecessary noisy inputs that could negatively impact performance while preserving the integrity of feature information. In addition, to gain in-depth information about the characteristics of wind turbine blades, a fusion module is suggested to implement the information fusion of wind turbine features. Specifically, each fan feature is mapped to a one-dimensional vector with the same length, and all one-dimensional vectors are concatenated to obtain a two-dimensional vector. And then, in the fusion module, the information fusion of the same characteristic variables in the different channels is realized through the Channel-mixing MLP, and the information fusion of different characteristic variables in the same channel is realized through the Column-mixing MLP. Finally, the fused feature vector is input into the Transformer for feature learning, which enhances the influence of important feature information and improves the model's anti-noise ability and classification accuracy. Extensive experiments were conducted on the wind turbine supervisory control and data acquisition (SCADA) data from a domestic wind field. The results show that compared with other state-of-the-art models, including XGBoost, LightGBM, TabNet, etc., the F1-score of proposed gated fusion based Transformer model can reach 0.9907, which is 0.4%–2.09% higher than the compared models. This method provides a more reliable approach for the condition detection and maintenance of fan blades in wind farms.

KEYWORDS

Crack detection; gated residual network; fusion; attention

1 Introduction

With the development of industry and technology, wind energy is widely used for its cleanness, environmental protection and sustainability [1]. Wind power technology has become a new technology growth point and entered a stage of rapid development [2]. However, due to the harsh working environment and the increasingly complex mechanical structure, wind turbines are prone to failure, which not only increases the maintenance cost of the wind power plant, but also seriously affects the power output [3,4].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Blades are one of the main components of wind turbines. Blades working safely and efficiently has an essential impact on the stable operation of the wind turbine. However, the harsh working environment, unceasing working, and the wear between the blades and other unit components will accumulate potential failures, quickly leading to cracks and damage to the wind turbine blades. According to the Global Wind Energy Council statistics during eight years of the wind farm operation, about 30% of the accidents that causing shutting down for more than 7 days are caused by blade faults [5]. Blade crack harms wind turbine shafting, tower cylinder and other unit components, and if not treated, they will lead to accidents [6]. The cost and time required for repairing blade crack are high [7,8]. Therefore, detecting the state of the blade, especially when the blade cracks under complex working conditions, can quickly and accurately judge the fault state of the blade, which is very important for ensuring overall efficiency, safety and reliability of the wind turbine, reducing the loss and improving the economic benefit [9,10].

The traditional fault detection methods on wind turbine blades include acoustic emission detection [11], infrared imaging detection [12], and abnormal vibration detection [13]. Although these traditional fault detection methods have the advantages of fast detection speed and strong real-time, however, they are usually unable to detect the faults of the wind turbine blades in the working state, and have extremely high requirements on the detection environment and equipment, making them unable to be widely used in practice.

In recent years, commonly used wind turbine blade fault detection methods can be classified as model-based and data-driven [14]. Empirical mode decomposition (EMD) and discrete mathematical model are often used in wind turbine blade crack fault detection by using the changes in natural frequencies [15,16]. By applying the residual to describe the deviation degree of each characteristic parameter value, and through progressive parameterization, a blade icing identification model is established [17]. However, in practical applications, due to the difficulty of modeling multiple couplings and unexpected disturbances of system parameters, model-based methods often fail, and fan blade fault detection methods based on data analysis have received more extensive attention. Machine learning based data analysis methods do not need complex physical model construction and can utilize SCADA (supervisory control and data acquisition, SCADA) system data [18]. These methods include support vector machine (SVM) [19,20], XGBoost, deep neural network (DNN), etc. DNN and other methods extract the information from SCADA data to realize the fault detection of fan blades [21]. Bayesian belief networks (BNN) were also adopted for fault detection of wind turbine blades [22,23]. At present, in the field of fault detection, the gradient boosting decision tree (GBDT) algorithm is one of the most widely used algorithms [24,25]. Meanwhile, in unsupervised learning research, a wavelet enhanced autoencoder model (WaveletAE) to identify wind turbine dysfunction by analyzing the multivariate time series monitored [26].

To increase the quality and reliability of data for better analysis and prediction, noise reduction in data mining aims to eliminate extraneous noise and outliers from the data. A prominent feature screening technique is principal component analysis (PCA), which can eliminate noise, reduce dimensionality, and improve model performance [27,28]. However, because PCA is a linear transformation technique, it can only be used with data that can be separated along linear axes. Tree models such as XGBoost and LightGBM are common techniques for reducing unnecessary characteristics in fault detection [29,30]. However, these techniques may overlook some crucial features and cannot ensure the integrity of variable information.

Wind turbine blade fault detection relies heavily on information fusion, which can enhance data quality, increase the accuracy and dependability of mining findings, and so better enable data mining

applications. Although XGBoost [31] and LightGBM [32] are frequently employed to handle data mining, the training data can introduce noise and outliers. TabNet [33] is a deep learning model for tabular data that can carry out operations on tabular data such feature selection, regression, and classification. Although TabNet is highly interpretable and performant, it is not suited for processing massive amounts of high-dimensional tabular data because of its great reliance on the quality of the input data. Although TabTransformer [34] performs well in terms of performance and interpretability, it is not suited for processing vast amounts of high-dimensional tabular data because it is very dependent on the quality of the input data. SAINT (Self-Attentive Interactions) [35] is a deep learning model for sequence data. It has large modeling capabilities and high accuracy, but it takes a long time to train and has a strong dependence on hyperparameters. In conclusion, classic data fusion models require a lot of training time and rely heavily on the accuracy of the input data.

To address these issues, this paper proposes a Gated Fusion based Transformer Model for crack detection on wind turbine blades by integrating GRN, a fusion module and a Transformer. GRN is a deep neural network structure that can be used for feature screening [36]. GRN combines a residual network and a gating mechanism, which can effectively extract useful features in input data and remove irrelevant or redundant features while preserving the integrity of feature information. This approach can improve the accuracy and generalization of the model. Meanwhile, this paper proposes a new information fusion module that is changed from the model MLP-Mixer [37]. The fusion module can extract features from the two-dimensional vectors of wind turbine blades to fuse different dimensions of information. Because the MLP in the module shares the weight in the mapping process of different columns and channels, the number of parameters is significantly reduced, and the complexity of the network is reduced.

This method takes the original SCADA data of wind turbine blades as the research object. It uses the SCADA data to detect the fault of wind turbine blade crack to ensure that maintenance personnel can timely and accurately judge whether the fault occurs and give early warning of the defect during the wind turbine operation. The contributions of this paper are summarized as follows:

1. This paper introduces Gated Residual Network (GRN). GRN, through the gating mechanism, can reduce any unnecessary noisy inputs that could negatively impact performance while preserving the integrity of feature information, enabling efficient information flow.
2. This paper proposes a fusion module that consists of Channel-mixing MLP and Column-mixing MLP. By using Channel-mixing MLP and Column-mixing MLP to map the columns acted on the data channel and map the rows acted on the data column of two-dimensional feature data, information fusion between the spatial and channel domains is realized. Because the MLP in the module shares the weight in the mapping process of different columns and channels, the number of parameters is significantly reduced, and the complexity of the network is reduced.
3. The experimental results show that the proposed method has a better detection effect and can better meet the actual fan maintenance requirements than the existing methods. At the same time, this paper also provides a new solution for table data processing.

The rest of this paper is organized as follows. Section 2 presents the available SCADA data used in this work. Section 3 introduces the model of this paper. Section 4 presents the experimental procedure and experiment setup. Section 5 introduces and analyzes the experimental results. Section 6 is the conclusion, including a summary of the method and future work.

2 Data Description

The dataset used in this paper comes from the operation data collected from 300 wind turbines published by State Power Investment Corporation for two consecutive months. SCADA system collects all kinds of information about wind turbines in real-time, including environmental, state, and control parameters. According to the wind turbine blade crack state data collected every 10 min by SCADA for wind farms, 48,339 sample files were recorded. Each file contained about 400 lines of data, and each sample had 75 features. The SCADA system collects two different types of data: discrete characteristic variables and continuous characteristic variables. Continuous variables monitor the operation and progress of each component in real time. Discrete variables indicate the operation status of each fan component. Among them, ‘state of reactive power control’, ‘current state value of wind turbine’, ‘current state value of wheel’, ‘state of yaw value’ and ‘yaw required value’ are discrete characteristic variables; the rest characteristics are continuous characteristic variables. An example of the SCADA dataset is provided in [Table 1](#). In addition, more detailed SCADA data can be seen in [Appendix A, Table A1](#).

Table 1: Sample of SCADA data for a limited selection of turbines and signals

WT	Wheel speed	Wheel angle	Speed measured value of overspeed sensor	State of yaw value	Blade angle		
					Blade angle 1	Blade angle 2	Blade angle 3
013	14.73	64	-14.57	1	7.1	7.16	7.14
013	14.51	64	-14.6	0	7	7.07	7.05
013	14.71	64	-14.65	0	7.4	7.47	7.45
013	14.84	64	-14.72	1	7.37	7.44	7.42
013	14.61	64	-14.57	0	7.43	7.5	7.48
013	14.67	64	-14.63	0	7.78	7.84	7.82
013	14.67	64	-14.59	0	7.54	7.61	7.58
013	14.81	64	-14.73	1	6.98	7.05	7.03
013	14.73	64	-14.6	1	7.54	7.61	7.59
013	14.66	64	-14.74	1	7.63	7.69	7.68

[Table 1](#) shows an example of data collected by a SCADA system. In the table, ‘WT’ uniquely identifies each wind turbine in the wind farm. ‘wheel speed’, ‘wheel angle’, and ‘blade angle’ are working condition parameters. ‘state of yaw value’ is a state parameter and is usually a discrete variable.

[Fig. 1](#) compares the kernel density estimation of wheel speed for an identical fan in normal and cracked blade states. Wheel speed is a working condition parameter. The figure shows maxima for both normal and cracked blade data value probability densities. Thus, it is impossible to distinguish between normal and defective fan blades using just one working condition eigenvalue.

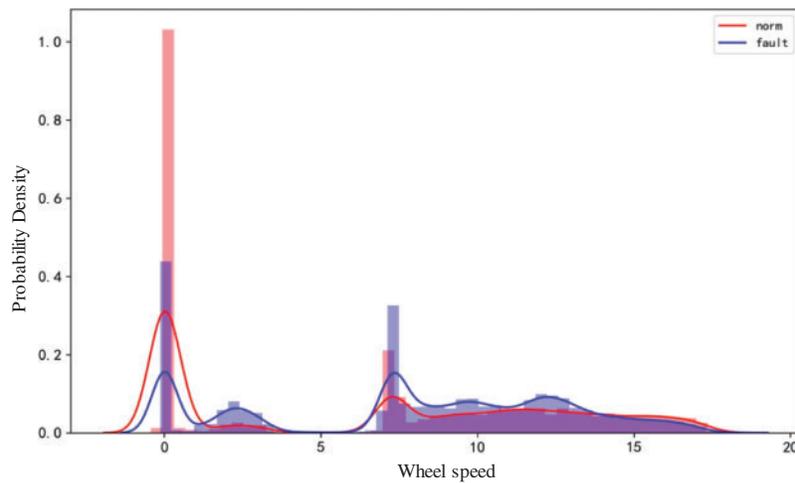


Figure 1: Wheel speed kernel density comparison chart

3 Model Structure

3.1 Overall Architecture

This paper presents gated fusion based Transformer model for crack detection on wind turbine blade. The proposed model mainly includes Gated Residual Network, fusion module and Transformer. The overall architecture of proposed model is shown in Fig. 2. First, each feature of the wind turbine blade is subjected to feature mapping, and the feature vectors are spliced to obtain a two-dimensional vector. Then, the gated residual neural network is used to effectively extract the features of the feature vectors and strengthen the expression of key features. Through the fusion module, the information exchange on different channels of the same feature and the same channel of different features can be realized. Use Transformer for feature learning, enhance the influence of important feature information and improve the anti-noise ability of the model. Finally, the state results of wind turbine blades are output through feature classification to achieve the purpose of crack fault detection.

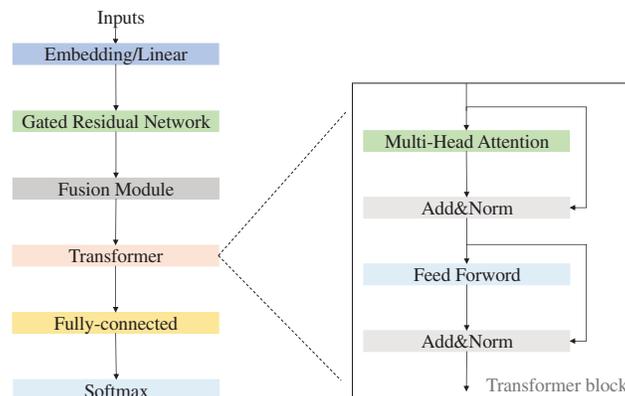


Figure 2: Structure of proposed model

3.2 Feature Mapping

The input of the model includes categorical feature variables and continuous feature variables. The proposed model uses feature embedding to embed each categorical feature variable into a parametric embedding of dimension d and map each continuous feature variable to a parametric embedding of dimension d through MLP. And then, the feature vectors are spliced to obtain a two-dimensional vector. Therefore, after feature mapping, the vector dimension input to GRN is $(m + n) * d$ if there are m discrete feature variables and n continuous feature variables.

3.3 Gated Residual Network

Due to subjective limitations in selecting input feature variables in commonly used fan blade feature screening methods, the integrity of variable information cannot be guaranteed. This paper uses GRN for feature screening. Gated Residual Network (GRN) enables efficient information flow with gating mechanism and skip connections. The framework of Gated Residual Network is shown in Fig. 3. The gating mechanism can reduce unnecessary noisy inputs that could negatively impact performance. Skip connections can provide adaptive depth and network complexity. In order to improve the performance of the model, a Dropout layer and a Swish activation function are simultaneously added to the network. The formula of skip connection is shown in Eq. (1).

$$H = F(x) + GLU(x) * x \quad (1)$$

where x is the input, $F(x)$ is a residual block, $GLU(x)$ is a gating unit, and $*$ represents the dot product operation.

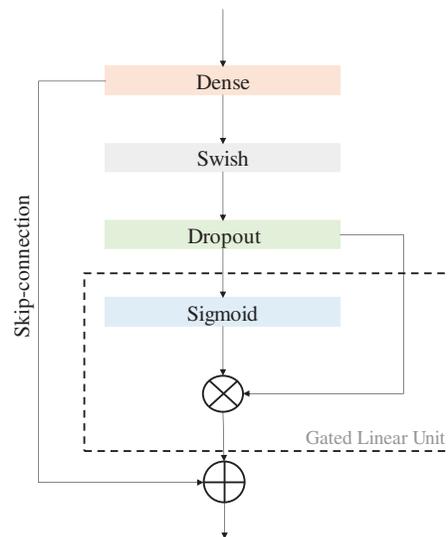


Figure 3: Structure of gated residual network

The formula of gating unit is shown in Eq. (2).

$$GLU(x) = x * sigmoid(x) \quad (2)$$

where x is the input, the sigmoid function is a typical non-linear activation function that transfers input values to continuous output values ranging from 0 to 1.

The Swish activation function is similar to both ReLU and Sigmoid activation functions. It combines the smoothness of Sigmoid with the nonlinear of ReLU. The formula for the Swish activation function is shown in Eq. (3).

$$F(x) = x * \text{sigmoid}(\beta * x) \quad (3)$$

where β is a hyperparameter, usually set to 1.

3.4 Fusion Module

The fusion module in this paper is modified from the MLP-Mixer model [37]. The MLP-Mixer network model simulates the convolution operation and attention mechanism using only the Multi-Layer Perceptron (MLP) architecture. The original MLP-Mixer first separates the image and then uses two MLP blocks to fuse data on each channel. Fusion module uses Channel-mixing MLP and Column-mixing MLP alternately to extract features from the two-dimensional vectors of wind turbine blades to fuse different dimensions of information. The framework of fusion module is shown in Fig. 4. The Channel-mixing MLP and Column-mixing MLP carry out mapping transformation of two-dimensional characteristic data in different channels of the same characteristic variables and the same channel of different characteristic variables, thus achieving fusion operations in the spatial domain and channel domain. To be specific, the feature vector is input into the MLP to obtain in-depth information about the same feature in different channels. Then, the processed feature vectors were transposed so that the channels and dimensions of the data were exchanged and then input into the MLP so that it could obtain in-depth information about different features in the same channel, and then the channels and dimensions were transposed. Because the MLP in the module shares the weight in the mapping process of different columns and channels, the number of parameters is significantly reduced, and the complexity of the network is reduced.

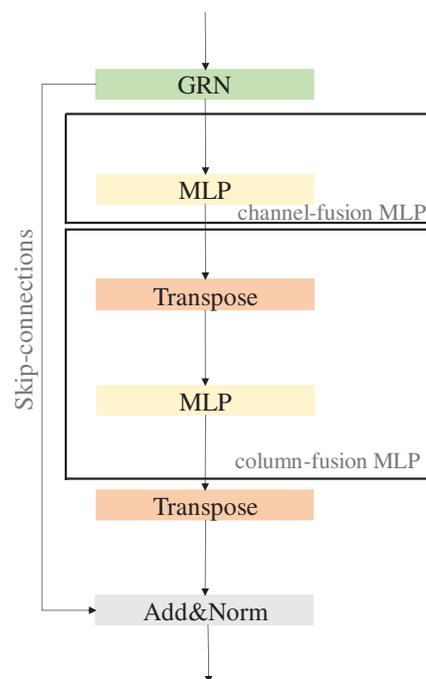


Figure 4: Structure of fusion module

4 Experiment Setup

In this section, the effectiveness of the proposed method will be evaluated through a case study of wind turbine blade crack fault detection. The experiment of the model in this paper is completed based on Tensorflow.

4.1 Overall Experimental Process

To provide a reliable basis for wind farm staff to monitor the status of fan blades and avoid more significant losses caused by blade crack, it is necessary to timely and accurately detect whether there is a crack fault in wind turbines. Therefore, the proposed wind turbine blade fault detection method is as follows.

4.1.1 Data Preprocessing

SCADA data should be cleaned to solve the problem of missing or unreasonable partial characteristic data caused by sensor errors, line errors and other issues.

4.1.2 Data Segmentation

Four-fold cross-validation was used to divide the cleaned data set into 4 subsamples, one of which was selected as the test set each time, the other 3 subsamples were chosen as the training set to conduct model training, and 10% of the training set was selected as the verification set, which was repeated for a total of 4 times, which was equivalent to using 4 kinds of data sets to train the model.

4.1.3 Model Training and Testing

The experimental flow chart is shown in Fig. 5. Appropriate evaluation criteria were used to evaluate the results of each experiment. Each model had a consequence. Finally, the mean value of the four marks of each model was calculated as the final experimental result.

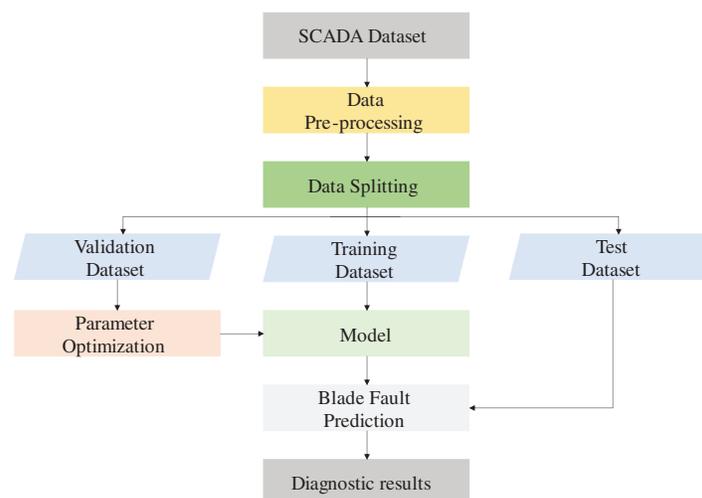


Figure 5: Flow chart of crack detection on wind turbine blade

4.2 Data Cleaning and Preprocessing

In real life, the database is easily affected by noise and some force majeure external factors, resulting in missing data and errors. Correct data processing can reduce the interference of erroneous data in the experiment and improve the efficiency and accuracy of data mining. In this experiment, the collected SCADA data were preprocessed as follows:

1. Delete data with integer row 0. Data with an integer value of 0 is obviously due to a sensor or wire error, so it needs to be deleted.
2. Delete the last three features where the data is null: 'Drive outputs torque 1', 'Drive outputs torque 2', 'Drive outputs torque 3'.
3. According to physics knowledge, the atmospheric pressure on the blade under the natural environment cannot be 0, so the row where the characteristic atmospheric pressure is 0 in the data set is deleted. In addition, this paper is to forecast the wind turbine in a working state, so the row where the Wheel Speed is less than 3 m/s is deleted.
4. The four-fold cross-validation was used to divide the cleaned data set into 4 subsamples. Each time, one of the subsamples was selected as the test set, and the other 3 subsamples were selected as the training set for model training. 10% of the training set was selected as the verification set, and each model repeated the experiment 4 times.

4.3 Performance Evaluation Index of Fault Diagnosis

The model evaluation index can reasonably measure the model's generalization ability and accurately explain the quality of the model. The prediction results of this model belong to the binary problem. Accuracy is the proportion of correctly classified samples in total samples. Precision refers to the proportion of the predicted positive samples to the predicted positive samples. Recall is the proportion of the predicted positive samples out of the actual positive samples. F1-score is mainly used to measure the accuracy of the binary model, taking into account the accuracy rate and recall of the classification model. Accuracy can intuitively see the prediction effect of the model, Precision and Recall can evaluate the prediction effect from different aspects, while F1-score combines Precision and Recall. So with Accuracy, Precision, Recall and F1-score index to evaluate the model is more reasonable.

4.4 Comparison Models

To verify the proposed model's effectiveness and accuracy in detecting wind turbine blade crack faults, it is compared with state-of-the-art methods such as XGBoost, LightGBM, LSTM, TabNet, TabTransformer, Variable Selection Network, and SAINT.

4.5 Hyperparameter

The selection of hyper-parameters is of great importance in modeling. For all model training, Adam is used for network optimization for its efficient computation and little memory. The cross entropy loss function is used in all models to train the network. Meanwhile, the initial learning rate of all neural network models is set as 0.01, if the val_loss of the model does not drop 3 times, the learning rate will be halved to help the model converge and make it easier to obtain the best solution. If the val_loss of the model does not drop 10 times, the model will stop early to reduce overfitting. The proposed model has Embedding_dim of 64, GRN_dim of 192, Num_transformer_blocks of 4, Num_heads of 4, Dropout_rate of 0.2, and Learning_rate is 0.01, Epoch is 100 and Batch_size is 256.

For the comparison models, XGBoost has Max_depth of 300, N_estimators of 500, Colsample_bytree of 0.8 and Subsample of 0.7; LightGBM has Num_leaves of 500, N_estimators of

50, Max_depth of 0.8 and Subsample of 0.7; LSTM has Num_input_layers of 256 and Number_hidden_layers of 2; VSN has Encoding_size of 192 and VSN_dim of 72; SAINT has Num_self-attention_heads of 8, Num_intersample_heads of 8 and Embedding_dim of 64; TabNet has N_d of 324, N_a of 256 and N_steps of 2; TabTransformer has Num_transformer_blocks of 4, Num_heads is 4 and Embedding_dim is 64.

5 Results and Discussions

5.1 Comparative Experiment

The proposed model is compared with comparison models in F1-score, Accuracy, Recall and Precision. The experimental results are the mean values of the results of 4-fold cross-validation, as shown in Table 2. The visualization of experimental results is shown in Fig. 6.

Table 2: Performance comparison between models

Model	F1-score	Accuracy	Recall	Precision
XGBoost	0.9835 ± 0.0019	0.9838 ± 0.0017	0.9779 ± 0.0035	0.9891 ± 0.0014
LightGBM	0.9869 ± 0.0006	0.9869 ± 0.0005	0.9815 ± 0.0008	0.9925 ± 0.0011
LSTM	0.9698 ± 0.0052	0.9695 ± 0.0044	0.9601 ± 0.0098	0.9797 ± 0.0046
TabNet	0.9740 ± 0.0047	0.9738 ± 0.0063	0.9653 ± 0.0052	0.9828 ± 0.0035
VSN	0.9827 ± 0.0029	0.9826 ± 0.0035	0.9757 ± 0.0022	0.9898 ± 0.0036
TabTransformer	0.9788 ± 0.0025	0.9787 ± 0.0023	0.9711 ± 0.0039	0.9867 ± 0.0022
SAINT	0.9853 ± 0.0035	0.9894 ± 0.0034	0.9844 ± 0.0043	0.9862 ± 0.0037
Proposed model	0.9907 ± 0.0043	0.9907 ± 0.0024	0.9853 ± 0.0045	0.9961 ± 0.0011

The result showed that compared with other seven algorithms, the gated fusion based Transformer model works best in F1-score which is around 0.9907, Accuracy which is around 0.9907, Recall which is around 0.9853 and Precision which is around 0.9961. The LSTM algorithm works worst in evaluation indicators, it reaches around 0.9698 in F1-score which is reduced by around 0.0209 compared with proposed model.

Analyzing the overall experimental results, it can be found that the traditional machine learning models such as XGBoost and LightGBM are better than the other deep learning models such as LSTM, TabNet, VSN, TabTransformer. And there is little difference between experimental results of LightGBM and SAINT. The F1-score of LightGBM is around 0.9896, which is increased by around 0.0042–0.0171 compared with LSTM, TabNet, VSN and TabTransformer. Compared with XGBoost and LightGBM, the F1-score value of the model used in this paper has increased by 0.0072 and 0.0038, respectively.

From the perspective of stability, it is clear that the fault detecting ability of machine learning models is more stable than other models including the gated fusion based Transformer model. Comparing proposed model and other deep learning models, the fluctuation of Accuracy and Precision range is too small. In conclusion, the Gated Fusion based Transformer Model proposed in this paper has high prediction accuracy and stability.

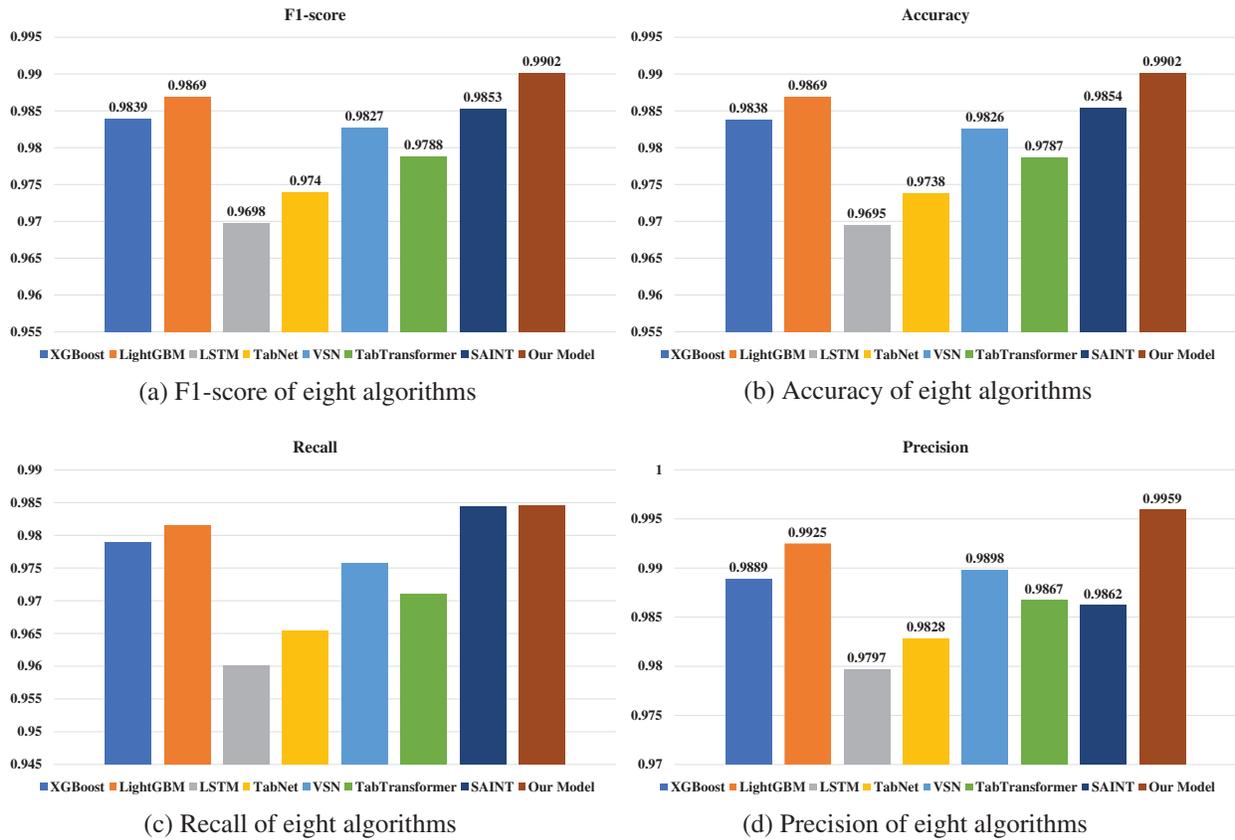


Figure 6: The values of evaluation indicators for models

5.2 Ablation Experiment

This paper also verifies the impact of each component of the modules including GRN, Fusion Model and Transformer on the fault detection model. The experimental results are shown in Table 3.

Table 3: Ablation on inclusion of submodules

Model	F1-score	Accuracy	Recall	Precision
Transformer	0.9889	0.9888	0.9827	0.9951
GRN based transformer	0.9894	0.9894	0.9832	0.9957
Fusion module based transformer	0.9894	0.9893	0.9835	0.9953
Proposed model	0.9907	0.9907	0.9853	0.9961

As shown in Table 3, the F1-score of Transformer is 0.9889. When GRN is included, the effect of the model is improved, and the F1-score is increased to 0.9894. Experiments prove the effectiveness of GRN. When the fusion module is included with the Transformer, the effect of the model is also improved, and the F1-score is increased to 0.9894. Finally, when the GRN and fusion modules are

both included with the Transformer, the proposed model achieves the best performance, with the F1-score of 0.9902. In the summary, the results prove the effectiveness of each submodules.

5.3 Discussion

This paper proposes a crack fault-detection strategy for high-precision wind blade crack detection. Experiments show that the proposed method outperforms conventional fault detection models. The F1-score of proposed model can reach 0.9907, which is 0.0040–0.0209 higher than traditional models. The accuracy of proposed model can also reach 0.9907, which is 0.0013–0.0212 higher than traditional models. Its recall can reach 0.9853, which is 0.0009–0.0252 higher than traditional models. Its precision can reach 0.9961, which is 0.0036–0.0164 higher than traditional models. These results demonstrate the superiority of the proposed model.

The ablation experiment reveal that the model's gated residual neural network, fusion module, and Transformer all contribute to superior experimental outcomes. GRN reduces unnecessary noisy inputs that could negatively impact performance while preserving the integrity of feature information. The fusion module fuses information between the spatial and channel domains to comprehensively obtain the in-depth information of fan features. The proposed model is suitable for structured data with both categorical and continuous feature variables. However, the gated fusion based Transformer model is less stable than XGBoost and LightGBM. The stability and generalization ability of the proposed model should be further improved.

6 Conclusions

Considering that the data collected by the SCADA system is complex and challenging to obtain the features related to the target information, a new state monitoring Model-Gated Fusion based Transformer Model is proposed in this paper, which is used for wind turbine blade crack fault detection. Gated residual network was used for effective feature extraction and enhanced expression of critical features. The fusion module proposed in this paper can map two-dimensional feature data in columns acted on the data channel and rows acted on the data column, realize the information fusion between the spatial domain and the channel domain, and obtain the in-depth information of fan features more comprehensively. Finally, the model can enhance the influence of important feature information and improve the detection accuracy of the model by Transformer. The experimental results show that the model yields state-of-the-art performance.

In summary, the method proposed in this paper outperforms other fault detection methods. The experiments used a dataset with single fault. Future research could study methods for detecting multiple types of faults in fan blade data.

Acknowledgement: The authors thank the support of Huaneng Offshore Wind Power Science and Technology Research Co., Ltd. and Yancheng Institute of Technology.

Funding Statement: This research was supported by the Jiangsu Provincial Key R&D Programme (BE2020034) and China Huaneng Group Science and Technology Project (HNKJ20-H72).

Author Contributions: Conceptualization, W.T. and C.L.; formal analysis, W.T. and C.L.; investigation, W.T. and B.Z.; writing, review and editing, W.T. and C.L.; supervision, C.L. and B.Z. All authors have read and agreed to the published version of the manuscript.

Availability of Data and Materials: The data utilized in this study is available upon request to the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Wymore, M. L., Van Dam, J. E., Ceylan, H., Qiao, D. J. (2015). A survey of health monitoring systems for wind turbines. *Renewable and Sustainable Energy Reviews*, 52, 976–990.
2. Chang, V., Chen, Y., Zhang, Z. J., Xu, Q. A., Baudier, P. (2021). The market challenge of wind turbine industry-renewable energy in PR China and Germany. *Technological Forecasting and Social Change*, 166, 120631.
3. Si, Y., Chen, Z., Zeng, W., Sun, J., Zhang, D. (2021). The influence of power-take-off control on the dynamic response and power output of combined semi-submersible floating wind turbine and point-absorber wave energy converters. *Ocean Engineering*, 227, 108835.
4. Rinaldi, G., Thies, P. R., Johannng, L. (2021). Current status and future trends in the operation and maintenance of offshore wind turbines: A review. *Energies*, 14(9), 2484.
5. Wang, Y., Tan, J., Qin, X., Hu, Z., Jing, Y. et al. (2021). Wind turbine blade fault diagnosis based on MFCC feature optimization strategy. *IOP Conference Series: Earth and Environmental Science*, vol. 675, 012074. Xiamen, China.
6. Oh, K. Y., Park, J. Y., Lee, J. S., Epureanu, B. I., Lee, J. K. (2015). A novel method and its field tests for monitoring and diagnosing blade health for wind turbines. *IEEE Transactions on Instrumentation and Measurement*, 64(6), 1726–1733.
7. Liu, Z., Wang, X., Zhang, L. (2020). Fault diagnosis of industrial wind turbine blade bearing using acoustic emission analysis. *IEEE Transactions on Instrumentation and Measurement*, 69(9), 6630–6639.
8. Katsaprakakis, D. A., Papadakis, N., Ntintakis, I. (2021). A comprehensive analysis of wind turbine blade damage. *Energies*, 14(18), 5974.
9. Gao, Z., Liu, X. (2021). An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. *Processes*, 9(2), 300.
10. Cho, S., Choi, M., Gao, Z., Moan, T. (2021). Fault detection and diagnosis of a blade pitch system in a floating wind turbine based on Kalman filters and artificial neural networks. *Renewable Energy*, 169, 1–13.
11. Bo, Z., Yanan, Z., Changzheng, C. (2017). Acoustic emission detection of fatigue cracks in wind turbine blades based on blind deconvolution separation. *Fatigue & Fracture of Engineering Materials & Structures*, 40(6), 959–970.
12. Wang, C., Gu, Y. (2022). Research on infrared nondestructive detection of small wind turbine blades. *Results in Engineering*, 15, 100570.
13. Yu, M., Fu, S., Gao, Y., Zheng, H., Xu, Y. (2018). Crack detection of fan blade based on natural frequencies. *International Journal of Rotating Machinery*, 2018, 1–13.
14. Sharma, V. B., Singh, K., Gupta, R., Joshi, A., Dubey, R. et al. (2021). Review of structural health monitoring techniques in pipeline and wind turbine industries. *Applied System Innovation*, 4(3), 59.
15. Wang, J., Liu, Z., Wang, Y., Wen, C., Wang, J. (2023). Extraction of strain characteristic signals from wind turbine blades based on EEMD-WT. *Energy Engineering*, 120(5), 1149–1162.
16. Sun, S., Wang, T., Yang, H., Chu, F. (2022). Damage identification of wind turbine blades using an adaptive method for compressive beamforming based on the generalized minimax-concave penalty function. *Renewable Energy*, 181, 59–70.
17. Dong, X., Gao, D., Li, J., Zhang, J., Zheng, K. (2020). Blades icing identification model of wind turbines based on SCADA data. *Renewable Energy*, 162, 575–586.

18. Liu, Y., Zheng, Y., Ma, Z., Wu, C. (2023). Wind turbine spindle operating state recognition and early warning driven by SCADA data. *Energy Engineering*, 120(5), 1223–1237.
19. Reder, M., Melero, J. J. (2018). A Bayesian approach for predicting wind turbine failures based on meteorological conditions. *Journal of Physics: Conference Series*, 1037(6), 062003.
20. Li, X., Cheng, B., Zhang, H., Zhang, X., Yun, Z. (2021). A hybrid model based on back-propagation neural network and optimized support vector machine with particle swarm algorithm for assessing blade icing on wind turbines. *Energy Engineering*, 118(6), 1869–1886.
21. Qiang, S., Lin, H., Yu, Z. (2016). Online faults diagnosis of wind turbine blades based on support vector machines. *2016 3rd International Conference on Systems and Informatics (ICSAI)*, pp. 247–250. Shanghai, China.
22. Santos, P., Villa, L. F., Reñones, A., Bustillo, A., Maudes, J. (2015). An SVM-based solution for fault detection in wind turbines. *Sensors*, 15(3), 5627–5648.
23. Pandit, R., Astolfi, D., Hong, J., Infield, D., Santos, M. (2023). SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. *Wind Engineering*, 47(2), 422–441.
24. Zhang, L., Liu, K., Wang, Y., Omariba, Z. B. (2018). Ice detection model of wind turbine blades based on random forest classifier. *Energies*, 11(10), 2548.
25. Tao, T., Liu, Y., Qiao, Y., Gao, L., Lu, J. et al. (2021). Wind turbine blade icing diagnosis using hybrid features and Stacked-XGBoost algorithm. *Renewable Energy*, 180, 1004–1013.
26. Yuan, B., Wang, C., Jiang, F., Long, M., Yu, P. S. et al. (1902). Waveletfnn: A deep time series classification model for wind turbine blade icing detection. arXiv preprint arXiv:1902.05625.
27. Wen, X., Xu, Z. (2021). Wind turbine fault diagnosis based on ReliefF-PCA and DNN. *Expert Systems with Applications*, 178, 115016.
28. Wang, Y., Ma, X., Qian, P. (2018). Wind turbine fault detection and identification through PCA-based optimal variable selection. *IEEE Transactions on Sustainable Energy*, 9(4), 1627–1635.
29. Wu, Z., Wang, X., Jiang, B. (2020). Fault diagnosis for wind turbines based on ReliefF and eXtreme gradient boosting. *Applied Sciences*, 10(9), 3258.
30. Liu, Y., Yang, C., Huang, K., Liu, W. (2023). A multi-factor selection and fusion method through the CNN-LSTM network for dynamic price forecasting. *Mathematics*, 11(5), 1132.
31. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y. et al. (2016). Xgboost: Extreme gradient boosting. *R Package Version 0.4-2*, 1(4), 1–4.
32. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W. et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.
33. Arik, S.Ö., Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 6679–6687. California, USA.
34. Huang, X., Khetan, A., Cvitkovic, M., Karnin, Z. (2020). Tabtransformer: Tabular data modeling using contextual embeddings. arXiv preprint arXiv: 2012.06678.
35. Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., Goldstein, T. (2021). Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. arXiv preprint arXiv: 2106.01342.
36. Lim, B., Arik, S.Ö., Loeff, N., Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764.
37. Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X. et al. (2021). MLP-Mixer: An all-MLP architecture for vision. *Advances in Neural Information Processing Systems*, vol. 34, pp. 24261–24272.

Appendix A**Table A1:** Feature names of SCADA data

No.	Parameter	No.	Parameter
1	Wheel speed	39	Generator stator temperature 5
2	Wheel angle	40	Generator stator temperature 6
3	Blade angle 1	41	Generator air temperature 1
4	Blade angle 2	42	Generator air temperature 2
5	Blade angle 3	43	Main bearing temperature 1
6	Motor current 1	44	Main bearing temperature 2
7	Motor current 2	45	Wheel temperature
8	Motor current 3	46	Wheel control cabinet temperature
9	Speed measured value of overspeed sensor	47	Room temperature
10	Mean value of yaw against wind in 5 seconds	48	Cabin temperature control cabinet
11	Vibration in the x direction	49	Inverter INU temperature
12	Vibration in the y direction	50	Inverter ISU temperature
13	Hydraulic brake pressure	51	INU RMIO temperature
14	Cabin weather station wind speed	52	Power estimation of variable propeller motor 1
15	The wind absolute value	53	Power estimation of variable propeller motor 2
16	State of reactive power control	54	Power estimation of variable propeller motor 3
17	The grid side current	55	Current state value of wind turbine
18	The grid side voltage	56	Current state value of wheel
19	Active power of inverter grid side	57	State of yaw value
20	Reactive power of inverter grid side	58	Yaw required value
21	Generator power of inverter generator side	59	Temperature of the blade battery box 1
22	Generator operating frequency	60	Temperature of the blade battery box 2
23	Generator current	61	Temperature of the blade battery box 3
24	Generator torque	62	Blade rotor motor temperature 1
25	Generator inlet temperature	63	Blade rotor motor temperature 2
26	Outlet temperature of the frequency converter	64	Blade rotor motor temperature 3
27	Frequency converter inlet pressure	65	Blade frequency converter box temperature 1
28	Frequency converter outlet pressure	66	Blade frequency converter box temperature 2
29	Generator power limiting values	67	Blade frequency converter box temperature 3
30	Reactive power set point	68	Blade supercapacitor voltage 1
31	Rated hub speed	69	Blade supercapacitor voltage 2
32	Wind tower ambient temperature	70	Blade supercapacitor voltage 3
33	Generator stator temperature 1	71	Drive thyristor temperature 1
34	Generator stator temperature 2	72	Drive thyristor temperature 2
35	Generator stator temperature 3	73	Drive thyristor temperature 3
36	Generator stator temperature 4	74	Atmospheric pressure
37	Torque 1	75	Torque 3
38	Torque 2		