Wind Turbine Blade Icing Diagnosis Using Convolutional LSTM-GRU With Improved African Vultures Optimization

WENHE CHEN\textsuperscript{1}, HANTING ZHOU\textsuperscript{1}, LONGSHENG CHENG\textsuperscript{1}, AND MIN XIA\textsuperscript{2} (Senior Member, IEEE)

\textsuperscript{1}School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
\textsuperscript{2}Department of Engineering, Lancaster University, LA1 4YW Lancaster, U.K.

CORRESPONDING AUTHORS: L. CHENG and M. XIA (e-mail: cheng_longsheng@njust.edu.cn; m.xia3@lancaster.ac.uk)

This work was supported by the China Scholarship Council (CSC) under Grant 202206840062.

ABSTRACT Wind farms are usually located in plateau mountains and northern coastal areas, bringing a high probability of blade icing. Blade icing even leads to blade cracks and turbine collapse. Traditional methods of blade icing diagnosis increase operating costs and have the potential risk of damaging the original mechanical structure. A data-driven model based on a novel convolutional recurrent neural network is proposed in this article. The method can effectively extract hidden features for accurate icing diagnosis. The hyperparameters of the proposed model are optimized by the improved African vultures optimization algorithm (IA VOA). To alleviate the critical data imbalance, the adaptive synthetic (ADASYN) is used to oversample the minority classes of icing status. In comparison to the state-of-the-art classification methods, the proposed method illustrates the outstanding effectiveness in blade icing diagnosis using the sensor data from supervisory control and data acquisition (SCADA) systems. The effectiveness analysis of variables, ablation study, and sensitivity analysis validates the performance of the proposed method.

INDEX TERMS Adaptive synthetic (ADASYN), convolutional (Conv) layer, icing diagnosis, stacked long short-term memory (LSTM)-gated recursive unit (GRU), wind turbine.

I. INTRODUCTION

WIND power, as renewable energy, can effectively replace fossil energy which has been widely used in recent years to solve the global energy crisis and the growing energy demand [1]. Wind farms are usually located in plateau mountains or northern coastal areas. In winter, wind turbine blades have a higher probability of icing due to the cold and wet environment. Icing can reduce the efficiency of wind turbine power generation and increase blade loading, leading to severe safety accidents, such as blade cracks and wind turbine collapse [2]. Therefore, an effective icing diagnosis of turbine blades is important. Much effort has been done to diagnose the icing status to minimize the losses.

Traditionally, wind turbine blade icing diagnosis includes direct and indirect methods [3]. The direct detection methods monitor the icing status of the blades, using auxiliary equipment, such as surveillance cameras, temperature and pressure sensors, and hyperspectral imaging systems. However, it can increase the mechanical complexity of wind turbines and installation, and maintenance costs, which leads to low feasibility and reliability. The indirect detection method discriminates the icing status, which relies on the operational data of the wind turbine, such as the comparison of output power, heated and unheated anemometers, dew point, and temperature. Its disadvantage is relying on expert experience and being insensitive to small amounts of ice [4].

To overcome the shortcomings of traditional methods, data-driven methods based on the monitoring signals of wind turbines have received more attention for icing diagnosis due to the low cost and fewer mechanical changes. The data-driven approaches can be divided into shallow machine learning and deep learning methods, which have been applied to the vital components of the wind turbine, such as gearboxes and bearings. In recent years, the supervisory control and data acquisition (SCADA) systems have been widely
installed and applied by wind farms, which provide the variables of the wind turbines, such as mechanical, electrical, and environmental [5]. Therefore, researchers apply variables of the wind turbines, such as mechanical, electrical, and environmental [5]. Therefore, researchers apply the hybrid model combined with data-driven methods and SCADA data for icing diagnosis. Shallow machine learning models establish the classification for diagnosis based on extracting the features of icing status, such as extreme gradient boosting (XGBoost) [6], kernel extreme learning machine (KELM) [7], and minority clustering synthetic oversampling technique [8]. The limitation of shallow machine learning methods is that achieving features usually requires domain engineering experience and is time consuming. The deep learning method does not require feature extraction relying on expert experience. It can automatically extract the features and try to model high-level representations using the SCADA data, such as bidirectional gated recurrent unit (BiGRU) [9] and temporal attention convolutional neural network (CNN) [10]. Its performance was more accurate than the shallow machine learning method. In addition, the appropriate hyperparameters can improve the diagnosis ability of the data-driven methods. Many researchers attempt to employ metaheuristic algorithms to optimize the hyperparameters, such as grey wolf optimizer (GWO) [11], particle swarm optimization (PSO) [12], and genetic algorithm (GA) [13]. However, the GWO algorithm is difficult to solve a large number of variables and escape local solutions when handling large-scale problems [14]. The PSO algorithm has poor effectiveness in high-dimensional and falls into premature convergence easily. And the GA algorithm does not handle high-dimensional problems well [15].

In this article, we develop a new deep learning framework for achieving a high-performance diagnosis of wind turbine blade icing. The convolutional (Conv) layer can efficiently extract features from wind turbine sensor data balanced by adaptive synthetic (ADASYN). Combining the advantages of long short-term memory (LSTM) and gated recursive units (GRU), the improved diagnosis method for deep recurrent neural networks (RNNs) is constructed with optimal hyperparameters through the African vultures optimization algorithm (AVOA) improved by the iterative chaotic map with infinite collapses (ICMICs). The main contributions of this article can be summarized as follows.

1) A new framework for icing diagnosis is proposed to implement the effective maintenance decision for wind turbine blades based on a data-driven model. The proposed method does not require additional auxiliary equipment and allows icing detection using SCADA data from the wind turbine. This method is cost effective and easily applicable.

2) The stacked LSTM-GRU is constructed to diagnose wind turbine blades icing accurately, which considers the temporal correlation between samples. The Conv layer is added to LSTM-GRU to extract primary features as input from the SCADA data.

3) The ratio of icing and nonicing samples in the wind turbine operation data is significantly imbalanced. The ADASYN is employed to oversample the icing samples of the minority class for addressing the data imbalance problem.

4) To automatically determine the hyperparameters in Conv-LSTM-GRU instead of relying on expert experience, improved AVOA (IAVOA) is employed to find the optimal solution. The automated process can achieve the optimal models for icing diagnosis with minimum training cost.

The remainder of this article is organized as follows. Section II introduces related work on wind turbine blade icing diagnosis, CNN, and stacked LSTM-GRU. Section III describes the proposed method in detail. Section IV evaluates the effectiveness of the proposed method. The conclusion and possible future work are given in Section V.

II. BACKGROUND

A. WIND TURBINE BLADE ICING DIAGNOSIS

Wind turbine blade icing diagnosis can be regarded as a time-series classification (TSC) task. In recent years, deep learning methods for solving TSC tasks have become popular. As a data-driven approach, deep learning can overcome the drawback of manual feature extraction and selection which requires human expertise and efforts, such as Stacked BiGRU [9], CNN-LSTM [16], and Conv-RNN [17]. These pieces of the literature concentrate on the structural design of the models for wind turbine blade icing diagnosis. However, the model for hyperparameters optimization and real-time icing diagnosis has not been fully explored.

B. CNN

The CNN is a variety of the feedforward neural network, which extracts and learns the hidden features from the input data [18]. CNN is composed of Conv layers, a batch normalization (BN) layer, and the ReLU layer. The convolution layer can efficiently perform temporal information and channel information fusion to completely extract the original data’s crucial hidden features by multiple convolution kernels. The convolution layer is formulated as

\[ h_t = \text{ReLU}(BN(W \ast x_t + b)) \]  

where \( W \) denotes the weight coefficient of the filters, \( x_t \) represents the input data, \( \ast \) is the convolution operation, \( b \) represents the bias parameter, ReLU denotes the active function, and \( h_t \) is the output data by the Conv layer.

C. STACKED LSTM-GRU

RNN is a self-feedback connected neurons network, which has good performance for TSC tasks. LSTM is a variant of RNN by introducing the gating mechanism for controlling the rate of accumulating information to avoid gradient vanishing. The LSTM units employ the three gates to replace the traditional hidden neurons, called input gate (\( i_t \), forget
FIGURE 1. Overall framework of the proposed icing diagnosis method.

III. PROPOSED METHOD
A. OVERALL FRAMEWORK
The proposed approach for wind turbine blade icing diagnosis is a data-driven method that needs no additional auxiliary equipment. The framework consists of three parts: 1) ADASYN for balancing data; 2) Conv-LSTM-GRU for icing diagnosis; and 3) IAVOA for hyperparameter optimization. To improve the reliability of the proposed model, the outliers and missing values need to be processed. Meanwhile, the nonicing status is much longer than the icing status. The imbalanced data can cause significant challenges to train a good model. Therefore, ADASYN is utilized to oversample the icing samples to balance the majority and minority classes. We employ the LSTM-GRU to diagnose the icing status considering the temporality between samples. Meanwhile, the Conv layer is added for extracting the primary features. Furthermore, the IAVOA is developed to optimize model hyperparameters, where ICMIC is used for chaotic initialization of the population, spiral search for expanding searching range, and Levy flight for saving computing costs. In the optimization process, we use the F1 score as the fitness function. Finally, the Softmax layer is set to output the icing probability of the wind turbine blade. The structure is shown in Fig. 1.

B. ADASYN FOR BALANCING DATA
ADASYN is applied as a sample synthesis algorithm by oversampling the minority class to address imbalanced data. The number of minority samples to be synthesized is determined by the number of majority samples in the K-nearest neighborhood [20]. In contrast to the synthetic minority oversampling technique (SMOTE), ADASYN concentrates on the data at the boundary between the minority class and the majority class which contributes to improving the accuracy and reliability of classification.

C. CONVOLUTIONAL LSTM-GRU
The proposed diagnosis model comprises the input, Conv, LSTM, GRU, and output layers, as shown in Fig. 2. The normalized historical SCADA data are imported into the Conv layer, which extracts the hidden features using convolutional kernels. Then, the feature data after dimensional reduction are imported into LSTM layers and GRU layers, training the network and learning features. Moreover, the dropout layer is utilized to reduce the occurrence of overfitting. We apply backpropagation through time (BPTT) to backpropagate the training error for updating the model parameters. A fully connected layer and Softmax layer are applied to produce the output. The fully connected layer acquires the evaluation score of each class by summing the weight of features. And the icing classification is completed after calculating...
the probability by the Softmax layer, formulated as
\[ \hat{y} = \text{Softmax}(z) = \text{Softmax} \left( W^T x + b \right) \] (7)
\[ \text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}}, j = 1, 2, \ldots, k \] (8)
where \( x \) denotes the input of the dense layer, \( W \) and \( b \) are the weight and bias, respectively, \( \hat{y} \) represents the probability of the classes, and \( K \) is the number of classes. Furthermore, the cross-entropy loss (CL) function is applied as a loss function to optimize the model, defined as
\[ \text{CL} = -y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}) \] (9)
where \( y \) is the label of the dataset, and the positive/negative samples are denoted as 1/0.

D. AVOA IMPROVED BY ICMIC
To enhance the performance of the proposed method, IAVOA is constructed to optimize the hyperparameters of Conv-LSTM-GRU. AVOA is a novel swarm intelligence optimization algorithm proposed recently, which simulates the feeding and flying process of the African vulture to obtain the solution to the optimization problem [21]. AVOA employs random variables to generate the initial population, with low ergodicity and individual nonuniform distribution. ICMIC map is a chaotic model with folding times wirelessly, which generates chaotic sequences in \((-1, 1)\). It has uniform ergodicity and fast convergence [22] compared with the logistic map and Tent map. IAVOA initializes the population by ICMIC map, increasing the diversity of the population to facilitate obtaining a better optimal global solution than AVOA.

1) IMPLEMENTATION PROCESS
Phase 1 (Chaotic Initialization Population): Assuming that there are \( n \) African vultures in the \( d \)-dimensional searched space. Then, the position of the \( i \)th vulture in the \( d \)-dimensional searched space is \( x_i = [x_{i1}, x_{i2}, \ldots, x_{id}] \), where \( i = 1, 2, \ldots, n \), \( j = 1, 2, \ldots, d \); \( x_{ij} \) is the position of the \( i \)th vulture in the \( j \)th dimension. The populations of IAVOA are initialized by the ICMIC mapping, formulated as
\[ x'_i = x_{ib} + (x_{ub} - x_{ib}) \cdot \frac{1 + z_i}{2} \] (10)
where \( x'_i \) denotes the chaotic initialized population; \( x_{ub} \) and \( x_{ib} \) denote the upper and lower bounds for each individual in each dimension, respectively; and \( z_i \) is the chaotic sequence using the ICMC mapping, expressed as
\[ z_i = \left\{ \begin{array}{ll} \sin(\alpha \pi / z_{n+1}) & \text{if } z_{n+1} \neq 0 \\ 1 \leq z_n \leq 1 & \text{otherwise} \end{array} \right. \] (11)
where \( \alpha \in (0, +\infty) \) is the constant value, which is helpful to obtain the good chaotic sequences, usually set as 0.6.

Phase 2 (Determining the Best Vulture in Any Group): After the initial population is obtained, we calculate the fitness of all solutions. The best solution is chosen as the best vulture in the first group, and the second best is chosen as the best vulture in the second group. Then, the other solutions shift toward the best solutions in the first and second groups, which recalculate the entire population in each interaction of fitness, expressed as
\[ R(i) = \begin{cases} \text{BestVal}_1, & \text{if } p_i = L_1 \\ \text{BestVal}_2, & \text{if } p_i = L_2 \end{cases} \] (12)
where \( L_1 \) and \( L_2 \) are the given parameters before the search operation, which are between 0 and 1, \( L_1 + L_2 = 1 \), and \( p_i \) is the probability of selected best solution.

Phase 3 (The Rate of Starvation of Vultures): Vultures often walk longer distances for seeking food when they are satisfied. However, they will search for food next to the stronger vultures and become aggressive when they are hungry. The rate of being satiated is utilized to transfer from the exploitation phase to the exploration phase. Since the rate of being satiated has a declining trend, the vultures randomly search for regions enhancing their satiety. \( F \) denotes the vultures that feel satiated, formulated as
\[ F = (2 \times \text{rand}_1 + 1) \times z \times \left( 1 - \frac{\text{inter}}{\text{inter}_{\text{max}}} \right) + t \] (13)
\[ t = h \times \left( \sin^\pi \left( \frac{\pi}{2} \times \frac{\text{inter}}{\text{inter}_{\text{max}}} \right) + \cos \left( \frac{\pi}{2} \times \frac{\text{inter}}{\text{inter}_{\text{max}}} \right) \right) - 1 \] (14)
where \( \text{inter} \) is the current iteration number, \( \text{inter}_{\text{max}} \) represents the total number of iterations, and \( z \) and \( h \) are random numbers between \((-1, 1)\) and \((-2, 2)\), respectively. If \( |F| \geq 1 \), vultures seek food in different areas and enter the exploration phase. If \( |F| < 1 \), vultures seek food in the solutions’ neighborhood and enter the exploitation phase.

Phase 4 (Exploration): This phase belongs to the exploration of the vultures. Vultures can check different random regions according to the two different strategies. And parameter \( P_1 \) is applied to select either strategy. Vulture’s position can be expressed as
\[ P(i + 1) = \begin{cases} R(i) - D(i) \times F & \text{if } P_1 \geq \text{rand}_{P_1} \\ R(i) - F + \text{rand}_2 \times ((u - l) + \text{rand}_3 + l) & \text{if } P_1 < \text{rand}_{P_1} \end{cases} \] (15)
where $P(i+1)$ denotes the vulture position vector in the next iteration, $D(i) = |X \times R(i) - P(i)|$, and $X$ denotes a coefficient vector that increases the random.

**Phase 5 (Exploitation):** When $0.5 < |F| < 1$, IAVOA belongs to the first stage of exploitation, which allows the vultures to operate rotational flight or siege strategy, expressed as

$$P(i + 1) = \begin{cases} D(i) \times (F + \text{rand}_4) - (R(i) - P(i)), & \text{if } P_2 \geq \text{rand}_2 \\ R(i) - (S_1 + S_2), & \text{if } P_2 < \text{rand}_2 \end{cases}$$

(16)

where rand$_4$ and rand$_2$ are the random numbers between 0 and 1; $P_2$ represents the given parameter, and $S_1$ and $S_2$ denote the rotational search state of the vultures. In the second stage of exploitation, the actions of the two vultures gather several types of vultures at the food source. When $|F|<0.5$, this algorithm phase is executed. rand$_3 \in (0, 1)$ belongs to the random number, P3 is a given parameter, and the process can be calculated as

$$P(i + 1) = \begin{cases} (A_1 + A_2)/2, & \text{if rand}_3 < P_3 \\ R(i) - \text{dis}(t) \times F \times \text{Levy}(d), & \text{if rand}_2 \geq P_3 \end{cases}$$

(17)

where dis$(t)$ represents the distance between the vulture and the best vultures in the two groups; $d$ denotes the variable dimension, $A_1$ and $A_2$ are the movement of vultures, and Levy$(\cdot)$ denotes the strategy of the Levy flight. This construction is suitable for tasks involving one piece of information at each time step.

2) **COMPLEXITY ANALYSIS**

Complexity analysis is a significant method for evaluating the performance of the algorithm. In AVOA, its computational complexity is determined by three phases, including initialization, fitness evaluation, and vultures updating. Assume that the number of vultures is $N$, the maximum number of iterations is $T$, and the number of dimensions is $D$. The computational complexity of the initialization phase is $O(N)$. The computational complexity of the update mechanism process of searching for the best position and updating the position vector of all formed vultures is $O(T \times N) + O(T \times N \times D)$. Therefore, the computational complexity of the AVAOA algorithm is $O(N \times (T + TD))$. IAVOA employs ICMIC instead of random variables to generate initialized populations. And its initialization, fitness, and vulture updating have the same complexity as AVOA. The computational complexity of IAVOA is $O(N \times (T + TD))$. Therefore, the proposed improved method does not reduce the solving efficiency of the algorithm. And the computational complexity of GWO is also $O(N \times (T + TD))$ [14]. Furthermore, the computational complexity of PSO is similar to $O(T \times D)$ [23]. The computational complexity of GA is $O(T \times N \times L)$, where $L$ is the length of the genotypes [24]. Therefore, we can observe that the computational complexity of IAVOA is the same as the GWO, and higher than the PSO and GA. However, the performance of obtaining the global optimal solution is higher than the other algorithms in the experiment.

**IV. EXPERIMENTAL EVALUATION**

**A. EXPERIMENTAL SETUP**

In this experiment, the icing data of wind turbine blades are obtained from the National Industrial Big Data Innovation Platform of China. The dataset contains nonicing and icing data, and each sample contains 26 variables, as listed in Table 1. We imported SCADA data of one wind turbine with a running time of 341.88 h. For the SCADA data, outliers are removed by the quartile method. And missing values are filled by the averaging method according to the consecutive values. The time points are sampled at 1-min interval in the SCADA system.

Fig. 3 illustrates the relationship between the power and the wind speed for the icing and nonicing statuses. The number of icing samples is 2841 and the number of nonicing samples is 39,621. And the imbalance rate (IR) of the data is 7.17%. 70% of the data is used as the training set and 30% as the testing set. The variables of sensors are the inputs, and the outputs are two categories, namely, icing and nonicing. The diagnosis model is to build the mapping between the input and output variables, indicated as $\{x_t\} = \{x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(26)}\} \rightarrow \{y_t\}$.

In the experiment, since the primary object in this article is icing status, the icing samples are considered positive samples, and the nonicing samples are considered negative samples. Then, this article uses recall, precision, and F1 score as evaluation indicators to evaluate the imbalance data, as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$

(18)
Recall = \frac{TP}{TP + FN} \quad (19)

F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)

where TP, FP, FN, and TN are true positive, false positive, false negative, and true negative, respectively.

**B. COMPARISON WITH OTHER DEEP LEARNING METHODS**

The ADASYN is applied to balance the unbalanced data for obtaining balanced training data. The number of positive samples and negative samples before oversampling is 891 and 13,466, respectively, and the IR is 93.38%. The number of positive samples after oversampling is 12,637, the number of negative samples is 13,466, and the IR is 6.16%, as shown in Fig. 4. The icing samples increase after oversampling using ADASYN, which keeps balance with the nonicing samples. We can observe some coincidences between the two classes. The primary reason is that ADASYN increases the number of training samples at the boundary between the minority and majority classes, which leads to the coincidence.

We apply the balanced data to demonstrate the proposed method for wind turbine icing diagnosis. In Conv-LSTM-GRU, its hyperparameters are determined by the IAVOA, where the initial learning rate (ILR), L2 Regularization (L2), and the numbers of LSTM layer (HN_LSTM) and GRU layer (HN_GRU) are searched in the ranges [0.001, 0.1], [0.0001, 0.001], [50, 200], and [50, 200], respectively. The optimal values of the four hyperparameters are 0.0015, 0.0003, 134, and 152, respectively. The total hyperparameter values and model structure settings are illustrated in Table 2.

Table 2. Parameter settings.

| Setting description | Setting and value |
|---------------------|-------------------|
| Numbers of units in input layer and output layer | 26, 2 |
| Number of Conv layers | 3 |
| Number of LSTM layer | 1 |
| Number of GRU layer | 1 |
| The kernel sizes of Conv layers | [5, 5], [8, 8], [3, 3] |
| The numbers of kernels in Conv layers | 64, 128, 64 |
| The number of hidden units in LSTM layer | 134 |
| The number of hidden units in GRU layer | 152 |
| Initial learning rate | 0.0015 |
| L2 Regularization value | 0.0003 |
| Learn rate drop factor | 0.1 |
| Number of iterations | 1000 |

Table 3. Evaluation indices in experiment 1.

| Models | Recall | Precision | F1  | Run time |
|--------|--------|-----------|-----|----------|
| Method1 (Proposed method) | 0.9896 | 0.9111 | 0.9487 | 102s |
| Method2 (LSTM) | 0.8978 | 0.7563 | 0.8210 | 95s |
| Method3 (GRU) | 0.9227 | 0.7773 | 0.8438 | 89s |
| Method4 (CNN) | 0.7781 | 0.8365 | 0.8062 | 92s |
| Method5 (XGboost) | 0.7531 | 0.7885 | 0.7704 | 98s |
| Method6 (SSAE) | 0.8105 | 0.7869 | 0.7985 | 115s |

Table 4. Hyperparameter setting using optimization algorithm.

| LAVOA | AVOA | GA | PSO | GWO | Default |
|-------|------|----|-----|-----|---------|
| ILR   | 0.0015 | 0.0026 | 0.0018 | 0.0001 | 0.0019 |
| L2    | 0.0003 | 0.0002 | 0.0003 | 0.0005 | 0.0002 |
| HN_LSTM | 134 | 155 | 120 | 73 | 200 |
| HN_GRU | 152 | 153 | 118 | 157 | 57 |

To validate the performance of the proposed method, five state-of-the-art methods are compared, including Method 2 (LSTM), Method 3 (GRU), Method 4 (CNN), Method 5 (XGboost), and Method 6 [stacked sparse autoencoder (SSAE)]. In LSTM and GRU, the structure of layers and neurons are set as “1-2-1” and “26-200-100-2,” respectively. CNN utilizes the same hyperparameters and structure described in Table 2. For hyperparameters of XGboost, max depth, gamma, and minimum child weight are 6, 0.1, and 3, respectively. In SSAE, the number of neurons and hidden layers is “100-200-100” and 3, respectively. For a fair comparison, the proposed method does not take into account the time of the optimization algorithm. Table 3 shows that the proposed model has almost the similar running time as other state-of-the-art methods but achieves higher accuracy.

**C. COMPARISON WITH OTHER OPTIMIZATION ALGORITHMS**

To evaluate the effectiveness of the proposed optimization algorithm, we compare it with four widely used algorithms including AVOA, GWO, PSO, and GA. To realize a fair comparison, these optimization algorithms are employed to optimize the hyperparameters of Conv-LSTM-GRU, with the same search range and fitness function illustrated in Section IV-B. The numbers of interaction and population in these optimization algorithms are 30 and 10, respectively. The searching processes and results of optimization algorithms are illustrated in Fig. 5 and Table 4.
As shown in Fig. 5, the proposed IAVOA is higher than other algorithms in convergence speed and fitness. IAVOA is less fitness than AVOA, GA, and GWO in the initial stage, due to applying the ICMIC to initialize the population for diversifying initial populations. From the 5th iteration, the fitness of IAVOA is higher than the other algorithms. Meanwhile, the iteration time obtaining the optimal solution for the optimization algorithm is taken as the run time. In Table 5, the recall, precision, and F1 score calculated by the comparative methods are smaller than the proposed method. And their run time is longer than the proposed method. The outcomes demonstrate that the proposed optimization algorithms perform well in obtaining the optimal global solution, which can achieve the highest accuracy under the minimum number of iterations.

D. EFFECTIVENESS ANALYSIS OF VARIABLES

Effectiveness analysis of variables is performed by random forest (RF). RF is a variable selection method that discerns its significance according to the score [25]. Fig. 6 illustrates the scores of all variables (Table 1). The top five scored variables are Yaw position, Generator speed, Active power, and Internal temperature of the nacelle, respectively. We apply the cumulative contribution rate (CCR) of LS “20%, 40%, 60%, 80%, and 100%” to select variables in different proportions for validating the effectiveness of the variable set. The comparative results are shown in Table 6, including Precision, Recall, and F1 score. It indicates that Conv and IAVOA can improve the performance of the stacked LSTM-GRU.

E. ABLATION STUDY AND SENSITIVITY ANALYSIS

The ablation study is implemented to validate the effectiveness of the proposed Conv module and the IAVOA module. To perform the ablation experiment, we remove the Conv and IAVOA, respectively, named _Conv and _IAVOA. Furthermore, the comparative results are illustrated in Table 7, including Precision, Recall, and F1 score. It indicates that Conv and IAVOA can improve the performance of the stacked LSTM-GRU.

Sensitivity analysis is performed to study the influence of important factors, including the Conv module and IR. In the Conv module, to acquire the impact of the number of Conv layers, we construct the different number of layers, setting as “1, 2, 3, 4.” The numbers of convolution kernels for the four networks are “64,” “64-128,” “64-128-64,” and “64-128-128-64,” respectively. To study the impact of convolution kernel size, the number of layers is set as 3, and we set the convolution kernel size, “[3, 3], [3, 3], [3, 3]” (Size1), “[5, 5], [5, 5], [5, 5]” (Size2), “[8, 8], [8, 8], [8, 8]” (Size3), and “[5, 5], [8, 8], [3, 3]” (Size4), respectively. As illustrated in Table 8, the highest recall, precision, and F1 score of the model are acquired when the number of Conv layers is set as 3. The worst performance of the proposed model is obtained when the number of Conv layers is set as 4. As illustrated in Table 9, the proposed model achieves the optimal recall, precision, and F1 score when the convolution kernel size is set as “[5, 5], [8, 8], [3, 3].” Furthermore, we randomly selected different proportions of icing samples according to different IRs for obtaining their influences. The IRs are set as “99%, 97%, 95%, 93.38%.” Table 10 shows
that the proposed model obtains the optimal recall, precision, and F1 score when the IR is set as 93.38%. Therefore, the proposed model achieves better performance when the IR decreases.

F. REAL-TIME ICING DIAGNOSIS

The real-time diagnosis framework is introduced to provide the online identification of icing status for wind turbine blades. We train the model in an offline form and perform the icing diagnosis in an online form using continuous sensor data. In this experiment, the hyperparameters and structures are the same as in the previous experiments. To illustrate the effectiveness of the proposed method, the results of real-time icing diagnosis compared with CNN are shown in Fig. 7. The light blue background denotes the icing period, which is marked by the engineers. The blue dashed line denotes the traditional threshold (0.5) for icing diagnosis. During the icing period, it is easily ensured that the proposed method obtains 100% accuracy. However, CNN only has about 75% accuracy and changes suddenly. In comparison to CNN, the proposed method can reduce false positives and maintenance costs.

V. CONCLUSION

To achieve an accurate diagnosis of blade icing for realizing condition-based maintenance of wind turbines, this article developed a novel diagnosis method based on Conv-LSTM-GRU-IAVOA. The Conv layer was established to extract features as input. The hyperparameters of the LSTM-GRU were optimized by IAVOA, which automatically achieved the optimal model. The proposed method was applied to accurately map the extracted features and blade conditions (icing and nonicing). To address the severe data imbalance, icing status as the minority class was oversampled by ADASYN. The experimental results illustrated the outstanding classification performance of the proposed method. Nevertheless, the proposed method still has some limitations. Some hyperparameters were still decided by experience, which could be further investigated. More sensors can be imported to enhance the accuracy of classification. The outliers in extreme conditions should be detected and processed to improve the diagnosis model.

REFERENCES

[1] Z. Ren, A. S. Verma, Y. Li, J. J. E. Teuwen, and Z. Jiang, “Offshore wind turbine operations and maintenance: A state-of-the-art review,” Renew. Sustain. Energy Rev., vol. 144, Jul. 2021, Art. no. 110886, doi: 10.1016/J.RSER.2021.110886.

[2] Z. Chen, J. M. Guerrero, and F. Blaabjerg, “A review of the state of the art of power electronics for wind turbines,” IEEE Trans. Power Electron., vol. 24, no. 8, pp. 1859–1875, Aug. 2009, doi: 10.1109/TPEL.2009.207082.

[3] A. Jiménez, F. P. García Márquez, V. B. Moraleda, and C. Q. Gómez Muñoz, “Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis,” Renew. Energy, vol. 132, pp. 1034–1048, Mar. 2019, doi: 10.1016/J.RENEENE.2018.08.050.

[4] A. G. Olabi, “State of the art on renewable and sustainable energy,” Energy, vol. 61, pp. 2–5, Nov. 2013, doi: 10.1016/J.ENERGY.2013.10.013.

[5] D. Astolfi, R. Pandit, L. Celesti, A. Lombardi, and L. Terzi, “SCADA data analysis for long-term wind turbine performance assessment: A case study,” Sustain. Energy Technol. Assess., vol. 52, Aug. 2022, Art. no. 102537, doi: 10.1016/J.SETA.2022.102537.

[6] T. Tao et al., “Wind turbine blade icing diagnosis using hybrid features and Stacked-XGBoost algorithm,” Renew. Energy, vol. 180, pp. 1004–1013, Dec. 2021, doi: 10.1016/J.RENEENE.2021.09.008.

[7] R. Tong, P. Li, X. Lang, J. Liang, and M. Cao, “A novel adaptive weighted kernel extreme learning machine algorithm and its application in wind turbine blade icing fault detection,” Measurement, vol. 185, Nov. 2021, Art. no. 110009, doi: 10.1016/J.MEASUREMENT.2021.110009.

[8] H. Yi, Q. Jiang, X. Yan, and B. Wang, “Imbalanced classification based on minority clustering synthetic minority oversampling technique with wind turbine fault detection application,” IEEE Trans. Ind. Informat., vol. 17, no. 9, pp. 5867–5875, Sep. 2021, doi: 10.1109/TII.2020.3046566.

[9] Y. Li et al., “Prediction of wind turbine blades icing based on feature Selection and 1D-CNN-SBiGRU,” Multimed Tools Appl., vol. 81, pp. 4365–4385, Jan. 2022, doi: 10.1007/s11042-021-11700-7.

[10] H. Cheng, F. Shi, M. Zhao, G. Li, H. Zhang, and S. Chen, “Temporal attention convolutional neural network for estimation of icing probability on wind turbine blades,” IEEE Trans. Ind. Electron., vol. 69, no. 6, pp. 6371–6380, Jun. 2022, doi: 10.1109/TIE.2021.30990702.

[11] A. Habibollahzade, I. Fakhari, S. Mohsenian, H. Aberoumand, and R. A. Taylor, “Multi-objective grey wolf optimization of solar chimneys based on an improved model incorporating a wind turbine power curve,” Energy Convers. Manag., vol. 239, Jul. 2021, Art. no. 114231, doi: 10.1016/J.ENCONMAN.2021.114231.

[12] J. Xu, W. Tan, and T. Li, “Predicting fan blade icing by using particle swarm optimization and support vector machine algorithm,” Comput. Elect. Eng., vol. 87, Oct. 2020, Art. no. 106751, doi: 10.1016/J.COMPELECENG.2020.106751.

[13] Z. Liu, X. Liu, K. Wang, Z. Liang, J. A. F. O. Correia, and A. M. P. de Jesus, “GA-BP neural network-based strain prediction in full-scale static testing of wind turbine blades,” vol. 12, no. 6, p. 1026, 2019, doi: 10.1016/J.COMPELECENG.2020.106751.

[14] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, “Grey wolf optimizer: A review of recent variants and applications,” Neural Comput. Appl., vol. 30, no. 2, pp. 413–435, 2018, doi: 10.1007/s00521-017-3272-5.

[15] I. Naruei and F. Keynia, “A new optimization method based on COOT bird natural life model,” Expert Syst. Appl., vol. 183, Nov. 2021, Art. no. 115352, doi: 10.1016/J.ESWA.2021.115352.
[16] J. Xiao, C. Li, B. Liu, J. Huang, and L. Xie, “Prediction of wind turbine blade icing fault based on selective deep ensemble model,” Knowl. Based Syst., vol. 242, Apr. 2022, Art. no. 108290, doi: 10.1016/J.KNOSYS.2022.108290.

[17] W. Tian, X. Cheng, G. Li, F. Shi, S. Chen, and H. Zhang, “A multilevel convolutional recurrent neural network for blade icing detection of wind turbine,” IEEE Sens. J., vol. 21, no. 18, pp. 20311–20323, Sep. 2021, doi: 10.1109/JSEN.2021.3093726.

[18] J. Gu et al., “Recent advances in convolutional neural networks,” Pattern Recognit., vol. 77, pp. 354–377, May 2018, doi: 10.1016/j.patcog.2017.10.013.

[19] M. Xia, H. Shao, X. Ma, and C. W. de Silva, “A stacked GRU-RNN-based approach for predicting renewable energy and electricity load for smart grid operation,” IEEE Trans Industr Inform, vol. 17, no. 10, pp. 7050–7059, Oct. 2021, doi: 10.1109/TII.2021.3056867.

[20] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN), 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.

[21] B. Abdollahzadeh, F. S. Gharehchopogh, and S. Mirjalili, “African vultures optimization algorithm: A new nature-inspired metaheuristic algorithm for global optimization problems.” Comput. Ind. Eng., vol. 158, Aug. 2021, Art. no. 107408., doi: 10.1016/J.CIE.2021.107408.

[22] J. Feng, J. Zhang, X. Zhu, and W. Lian, “A novel chaos optimization algorithm,” Multimed Tools Appl., vol. 76, no. 16, pp. 17405–17436, 2017, doi: 10.1007/s11042-016-3907-z.

[23] X. Zhang, D. Zou, and X. Shen, “A novel simple particle swarm optimization algorithm for global optimization.” Mathematics, vol. 6, no. 12, p. 287, Nov. 2018, doi: 10.3390/math6120287.

[24] B. Rylander, “Computational complexity and the genetic algorithm,” Ph.D. dissertation, College Grad. Stud., Univ. Idaho, Moscow, ID, USA, 2001.

[25] C. Lou and X. Li, “Unsupervised fault detection based on laplacian score and TEDA,” in Proc. IEEE 7th Data Driven Control Learn. Syst. Conf. (DDCLS), 2018, pp. 267–270, doi: 10.1109/DDCLS.2018.8515956.

HANTING ZHOU received the B.S. and M.S. degrees from China Jiliang University, Hangzhou, China, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree in machinery management science and engineering with the Nanjing University of Science and Technology, Nanjing, China.

Her research interests include smart manufacturing, signal processing, machine diagnosis and prognostics, and trustworthy AI.

LONGSHENG CHENG received the M.S. degree from the East China Institute of Technology, Nanchang, China, in 1988, and the Ph.D. degree from the Nanjing University of Science and Technology, Nanjing, China, in 1998.

He is currently a Professor of Management Sciences and Applied Statistics with the School of Economics and Management, Nanjing University of Science and Technology. His current research interests include prognostic and health monitoring, machine learning, quality engineering, and data mining.

MIN XIA (Senior Member, IEEE) received the B.S. degree from Southeast University, Nanjing, China, in 2009, the M.S. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree from the University of British Columbia, Vancouver, BC, Canada, in 2017.

He is currently a Lecturer with the Department of Engineering, Lancaster University, Lancaster, U.K. His research interests include smart manufacturing, machine diagnostics and prognostics, deep neural networks, wireless sensor network, and sensor fusion.