Xu, Z, Mei, X, Wang, X, Yue, M, Jin, J, Yang, Y and Li, C

Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors

Abstract: In order to solve the problems of insufficient extrapolation of intelligent models for the fault diagnosis of bearings in real wind turbines, this study has developed a multi-scale convolutional neural network with bidirectional long short term memory (MSCNN-BiLSTM) model for improving the generalization abilities under complex working and testing environments. A weighted majority voting rule has been proposed to fuse the information from multi-sensors for improving the extrapolation of multisensory diagnosis. The superiority of the MSCNN-BiLSTM model is examined through experimental data. The results indicate that the MSCNN-BiLSTM model has 97.12% mean F1 score, which is higher than existing advanced methods. Real wind turbine dataset and an experimental dataset are used to to demonstrate the effectiveness of the weighted majority voting rule for multisensory diagnosis. The results present that the diagnosis result of the MSCNN-BiLSTM model with weighted majority voting rule is higher respectively 1.32% and 5.7 % than the model with traditional majority voting or fusion of multisensory information in feature-level.

Keyword: Bearing; Wind turbine; convolutional neural network; fault diagnosis; information fusion

Nomenclature

MSCNN-BiLSTM	Multi-scale convolutional neural network with bidirection long short term memory
TICNN	Convolution neural networks with training interference
MSCNN-GRU	Multi-scale convolutional neural network with Gate Recurrent Unit
MS-CNN	Multi-scale convolutional neural network
MC-CNN	multi-scale cascade convolutional neural network
CNN	Convolutional neural network
GRU	Gate Recurrent Unit
EMD	Empirical mode decomposition
LSTM	Long short term memory
MCCNN-LSTM	Multi-convolution convolutional neural network with long short term memory
MA-CNN	Multi-head attention convolutional neural network
MSCNN	Multi-scale convolutional neural network
C-CNN	Parallel Convolution Layers with Multi-Scale Kernels
SNR	Signal-Noise-Ratio
WT	Wind turbine
LR	Logistic regression
Conv	Convolutional layer

BN	Batch Normalization
ReLU	Rectified Linear Unit
$y^{l(i,j)}$	the dot product of kernel
W	represents the width of the kernel
$\mathbf{K}_i^l(j')$	the j^{th} weight of kernel l.
$z^{l(i,j)}$	the output of one neuron
μ	the mean of $y^{l(i,j)}$
σ^2	the variance of $y^{l(i,j)}$,
ε	a small constant
$\gamma^{l(i)}$	the scale to be learned
$\beta^{l(i)}$	the shift parameters to be learned
$a^{l(i,j)}$	the activation of $z^{l(i,j)}$
$y_j^s(j)$	the output of the $x(i)$ processed by IMS procedure with the interference
$O_i(k)$	the k^{th} output feature
φ_i	the output of fully connected layer
α_i	The feature weight of each scale
g_t	Input gate
q_t	Output gate
f_t	Forget gate
c_t	Cell state
NREL	National Renewable Energy Laboratory
DAQ	Data Acquisition system

1. Introduction

Nowadays, various countries have paid more and more attention on the issues about energy security and ecological environment [1].The wind turbines (WT), as one of the most important renewable power productions, are developing rapidly in both terms of installed capacities and sizes because that vigorously developing clean renewable energy has become the universal consensus and concerted action of the international community to promote the transformation of the energy structure and respond to climatic variation [2]. Bearings are the key mechanical parts in a WT's transmission, the health conditions of which determine the power generation efficiency and stable operation of a WT. Therefore, diagnosis and monitoring for bearings in the WTs are necessary for reducing their maintenance costs and delaying service life [3].

On the one hand, in benefited from the development of deep learning techniques, a lots of neural network-based methods for maintenance and diagnosis have good graces in the age of digital information industry [4, 5, 6]. In this kinds of neural network-based diagnosis method studies, whether a diagnosis model is developed based on convolutional neural network [7], long short term memory [8] or adversarial network [9], the research points are the structure of network and the construction of data input [10]. But wind-induced vibration of wind turbine leads to complex operating environment of wind turbine [11-12], which leads to become difficulty for neural network-based fault diagnosis model. On the other hand, collaborative maintenance for multisensory diagnosis has become the research hotspot with the advent of the Industry 4.0. Single model performance and information fusion strategy all affect diagnostic results.

In order to improve the performance of a diagnostic model, images and raw vibration signals are used as the inputs for training a neural network-based model. Wang *et al*. [13] employed the wavelet spectrogram with a size of 32×32 as the input. The spectrogram, which is based on a 2-D CNN model, was adopted to identify different working states of the rotor systems. In their study, using different spare convolution neural network increased 5% than using ReLU network. Similarly, Chen *et al*. [14] used the continuous wavelet transform to gain representation images and then imported the images into a 2-D CNN model to address the fault diagnosis. Their model had 99.83% in their test experimental dataset. The difference between their study and Wang's study was that the classifier used in Chen's study is the extreme learning machine, leading to a higher performance under a fault diagnosis task for the rolling bearings. Considering this kinds of input data in the form of images will cause the loss of effective information, Jiang *et al.* [15], using 1-D vibration signals as the input data, proposed a diagnosis model based on the multi-scale convolutional neural network (MS-CNN) to diagnose gearbox faults of a wind turbine. The results indicated that the time scale of the MS-CNN model has a significant impact on the diagnosis effect of the model and got 98.53% on their experimental dataset. Zhao *et al.* [16] used 1-D vibration signal as the input of the proposed normalized CNN for an intelligent fault diagnosis of rolling bearings. The results show that the normalized CNN model has a better extrapolation ability by 98.50% than a traditional CNN model. Wang *et al.* [17] used 1-D CNN-based network to examine ten groups of bearings to validate its reliability. The results showed that the diagnostic performance of the model under variable conditions was improved to 99.93% because more fault information was considered. Wei *et al.* [18] adopted 1-D raw vibration signals of rolling bearing as the input of a deep CNN to simultaneously achieve feature extraction and classification. Huang *et al*. [19] convoluted the 1-D vibration signals by different kernel sizes to obtain different resolutions in frequency domain, which was introduced in to a CNN-based model to develop a multi-scale CNN diagnosis model to address the fault identification of bearings, which had 83.2% diagnosis result on their experimental dataset. Considering the contents of the fault information, Zhao *et al*. [20] proposed a bi-directional LSTM framework to monitor machine health. Lu *et al*. [21] used the LSTM with the deep neural network to address fault diagnosis at the beginning of failures. In summary, using raw vibration signals as the data set to train a neural network-based model for fault diagnosis are more robust than using images. Considering multi-scale information and the potential semantics relationships of fault information are helpful to improve a model's extrapolation performance. Therefore, combining the advantages of the above studies for establishing a single sensor model, the first motivation in this paper is to design the Multi-Scale CNNBiLSTM network for considering both multi-scale information capability and context association of fault information.

In the studies of multisensory information fusion strategies to machine diagnosis for collaborative maintenance, Jing et al. [22] and Azamfar et al [23] directly fused raw signals from multiple sensors as a multi-signals and used a CNN to extract advanced features for gearbox fault diagnosis. The above studies are based on the signal fusion level to process the information collected by multiple sensors. Although there are little loss through the date-level fusion, big data and noisy make those are not easy to achieve truly engineering. As an alternative, Chen et al. [24] and Liu et al. [25] first constructed a multi-sensor features then realize information fusion in the feature-level to finally diagnose fault. However, those kinds of feature fusions in the feature-level have better interpretabilities when facing the same category of information fusion. If the advanced features are derived from different information sources, the interpretability will be not strong enough. Therefore, the realization of information fusion in the decision-making level is a relatively suitable choice to address multi-sensor fault diagnosis for wind turbine maintenance [26]. Therefore, the second motivation in this paper is to design a weighted voting rule based on Genetic Algorithm (GA) for multisensory fault diagnosis. The disadvantages of CNN-based fault model, RNN-based model and multi-sensor fault diagnosis method are summarized in Table 1.

Table 1: Brief compared of diagnosis method

Methods	Advantages	Disadvantages	References
CNN-based model	End-to-end feature extraction Fast calculation	Ignoring the temporal correlation of fault features	15，19，40，41，42
RNN-based model	Considering the semantics of the fault features	Large amount of calculation	38
CNN-RNN-based model	End-to-end feature extraction Considering the semantics of the advanced fault features	Single feature extraction The order of advanced features is not considered	39，43
Multisensory diagnosis	Consider multiple sources of information	Strategy of information fusion is not considered	36，30

In order to improve the generalization abilities of a neural network-based model for fault diagnosis and fusing the diagnostic results from multiple sensors in a suitable way to increase the final diagnostic accuracies and robustness for wind turbine bearing maintenance. The Multi-Scale CNNBiLSTM model, based on multi-scale coarse-grained procedure algorithm, convolutional neural network and Bidirectional long short memory network, has been developed in this paper to capture multi-scale time information and associate fault semantic information. A weighted majority voting method based on Genetic Algorithm is proposed to fuse the diagnostic results corresponding to every sensor for improving robustness of the diagnostic method. The proposed Multi-Scale CNNBiLSTM model is examined through comparison with experimental data of noise and variable loading scenarios to verify its reliability and superiority in real wind turbine. The originalities and main contributions of this study are summarized as follows.

(1) The Multi-Scale CNNBiLSTM model, based on multi-scale coarse-grained procedure algorithm, convolutional neural network and Bidirectional long short memory network, has been developed in this paper to capture multi-scale time information and associate fault semantic information for improving the performance of a single model.

(2) An end-to-end intelligent diagnosis framework based on the Multi-Scale CNNBiLSTM model is developed to realize fault diagnosis of a rolling bearing, which is capable of directly operating on the measured raw signals without any manual modifications.

(3) A weighted majority voting method based on genetic algorithm has been proposed to fuse the diagnostic results of different sensors in decision-making level, which has better information fusion interpretation.

The remaining parts of the paper are organized as follows. The development of the Multi-Scale CNNBiLSTM framework is presented in Section 2. The experimental data of a test experimental data and the evaluation index is presented in Section 3. The validation and discussion of the Multi-Scale CNNBiLSTM model under various working conditions are presented in Section 4. Conclusions are presented in Section 5.

2. Methodologies about Multi-Scale CNNBiLSTM

2.1 Multi-scale extraction

A multi-scale coarse-grained process has been developed and implemented into a multi-scale feature extraction layer to extract more information from raw signals with multiple time scales [15]. However, the multi-scale layer in reference [15] adopted the non-continuous sampling when capturing the multi-scale information, which leads to omission of some inherent information.

The processing for calculating the traditional multi-scale coarse-grained procedure, is based on a given time series, $x_i : 1 \le i \le n$ and the coarse-grained time series as the time scale factor of τ in order to calculate a sub-signal y_j^τ through Eq. (1).

$$y_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, 1 \le j \le (n/\tau) \qquad (1)$$

where $\tau = 1, 2, 3, \ldots$ is the time scale factor. The length of the sub-signal y_j^τ is (n/τ).

An illustration of the traditional multi-scale operation for a time scale factor τ=3 is presented in Figure 1.

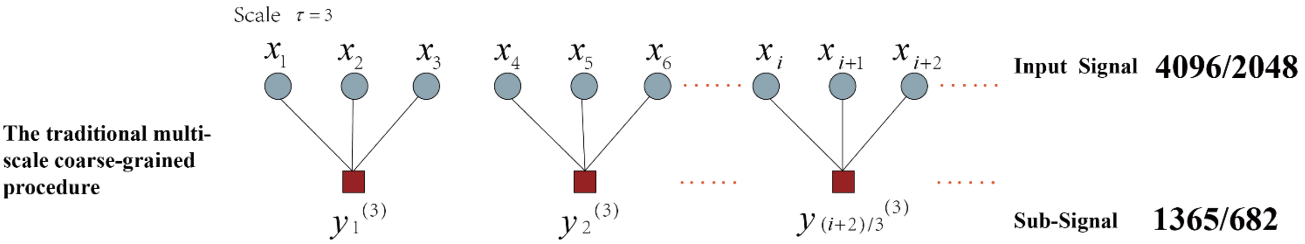


Figure 1: The traditional multi-scale coarse-grained operation

As shown in Figure 1, the length of the sub-signal decreases exponentially with increase in the time scale factor, which leads to its inability to perform the convolution process in a very deep convolution layer. More importantly, some useful information of the fault representation will not be captured due to discontinuity in the operation of a traditional MS coarse-grained procedure.

In order to solve the shortcomings of the traditional MS operation, a novel multi-scale coarse-grained procedure is developed and presented in this section. Figure 2 shows the continuous multi-scale coarse-grained procedure when the time scale factor τ is 3. The sub-signal z_j^n obtained by the CMS operation at any time scale factor τ is calculated by Eq (2).

$$z_j^{n-\tau} = \left\{ \frac{1}{\tau} \sum_{i=(j-1)+1}^{(j-1)+1+\tau} x(i),\ j \in [1, n\text{-}\tau], \tau \geq 2, z_{n\text{-}\tau}^{n} = \mathbf{0},\ j \in [n\text{-}\tau, n] \right. \tag{2}$$

where $\tau = 1, 2, 3, \ldots$ is the time scale factor. The length of the sub-signal z_j^n is n.

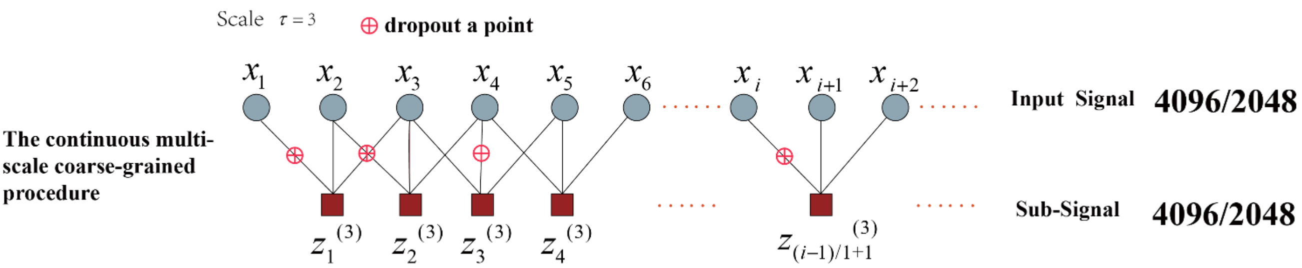


Figure 2: The continuous multi-scale coarse-grained operation

Compared with Figure 1, the length of the sub-signal processed by the CMS procedure will not decrease with time scale factor τ, which makes the CMS-based model easier to be maintained. In addition, some data points are randomly discarded using the dropout technology in the coarse-grained extraction process, in order to avoid the overfitting of data when training a model and to improve its robustness. Thus, the output z is given by Eq. (3).

$$\begin{cases} p \sim Uniform(0.1 \sim 0.2) \\ \mathrm{r}_i^1(k) \sim Bernoulli(p) \\ z_j^{\tau}(j) = \mathrm{r}_i^1 \cdot K_i^1 \cdot x_i \end{cases} \tag{3}$$

where ($\cdot$) represents the element-wise product, when the dropout rate p obeys the uniform distribution $U(0.1,0.2)$; $\mathrm{r}_i^1(k)$ follows the Bernoulli distribution, which is used to determine whether the k^{th} element in the i^{th} frame of the convolutional K_i^1 is dropped or not. $z_j^{\tau}(j)$ is the output of x_i processed by the CMS procedure with an interference in every batch training.

2.2 Feature learning layer

The feature learning layer consists of parallels of 1D CNNs that extract representation features from the sub-signals. Generally, a CNN structure is mainly composed of various pairs of convolutional layers and pooling layers. The activation function is used to realize the linear separation of the high-dimensional features after the convolution operations. $\mathbf{K}_i^l$ is the i^{th} filter in layer l, and $\mathbf{X}^{l(\mathrm{R}^j)}$ is j^{th} local area in the convolutional layer l. The convolutional process is given as follows:

$$y^{l(i,j)} = \mathbf{K}_i^l \cdot \mathbf{X}^{l(\mathrm{R}^j)} = \sum_{j'=0}^{W} \mathbf{K}_i^l(j')\mathbf{X}^{l(j+j')} \tag{4}$$

where $y^{l(i,j)}$ denotes the dot product of kernel and the local area. W represents the width of the kernel. $\mathbf{K}_i^l(j')$ is the j^{th} weight of kernel *l*.

In order to enhance the non-linear expression ability of the input signal and to more easily identify the learned features, the ReLU activation function is added after the convolutional layer. The formula for the ReLU is given in Eq. (8):

$$a^{l(i,j)} = f(z^{l(i,j)}) = \max\{0, z^{l(i,j)}\} \tag{5}$$

where $z^{l(i,j)}$ is the output array of the Batch Normalization (BN) and $a^{l(i,j)}$ is the activation of $z^{l(i,j)}$.

In order to efficiently accelerate the network training and to avoid the problem of gradient disappearance caused by activation function, the BN technique is introduced before the pooling operation. The *n*-dimensional array $\mathbf{y}^l = (y^{l(1)}, y^{l(2)}, \ldots, y^{l(n)})$ to the l^{th} BN layer is represented as $\mathbf{y}^{l(i)} = (y^{l(i,1)}, y^{l(i,2)}, \ldots, y^{l(i,n)})$ and $\mathbf{y}^{l(i)} = y^{l(i)} = y^{l(i,1)}$ when the BN layer is placed after the convolutional layer and fully connected layer, respectively. The formula for the BN operation is presented as follows:

$$\hat{y}^{l(i,j)} = \frac{y^{l(i,j)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}, z^{l(i,j)} = \gamma^{l(i)}\hat{y}^{l(i,j)} + \beta^{l(i)} \tag{6}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y^{l(i,j)} \tag{7}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (y^{l(i,j)} - \mu)^2 \tag{8}$$

where $z^{l(i,j)}$ is the output of one neuron. μ and σ^2 are the mean and variance of $y^{l(i,j)}$, respectively. ε is a small constant introduced to prevent the calculation from being invalid when the variance is 0. $\gamma^{l(i)}$ and $\beta^{l(i)}$ are respectively the scale and shift parameters to be learned.

The pooling layer is also called the down-sampling layer. The most common pooling techniques include average pooling and maximum pooling.The maximum pooling is chosen in this research, and it is presented in Eq (9).

$$p^{l(i,j)} = \max_{(j-1)W+1 \le t \le jW} \{a^{l(i,t)}\} \tag{9}$$

where $a^{l(i,t)}$ is the value of the t^{th} neuron in the i^{th} framework of layer l; W is the width of pooling size; $p^{l(i,j)}$ is the corresponding value of the neuron in layer l of the pooling, and $t \in [(j-1)W+1, jW]$.

2.3 The BiLSTM layer

The LSTM, proposed by Hochreiter *et al*. [27], is a variant of the Recurrent Neural Network (RNN). Using a standard RNN model [28] to calculate a given sequence $z_m = (z_1, z_2, z_3, \ldots, z_m)$ that is obtained by the CMSCNN layer for obtaining a hidden sequence $h = (h_1, h_2, \ldots, h_m)$ and an output sequence $Z_{\mathrm{m}} = (Z_1, Z_2, \ldots, Z_m)$. In order to overcome the shortcoming of the LSTM, thus in this study, BiLSTM is used to consider the semantic relevance of the information from both of the forward and backward of advanced features, which are represented as Eq. (10). The forward and backward information are fused into the fully connected layer and softmax function to calculate the probabilities of each failure.

$$Z = [\mathbf{Z}_f, \mathbf{Z}_b] \tag{10}$$

Where $\mathbf{Z}_f$ is the forward features; $\mathbf{Z}_b$ is the backward features.

The forward and backward features calculations are similarity. Take the forward features as an example, the calculation of Z_t is shown below, which also is the LSTM.

$$h_t = f_a(W_{zh}x_t + W_{hh}h_{t-1} + b_h) \tag{11}$$

$$Z_t = W_{zh}h_t + b_z \tag{12}$$

where W represents the weight coefficient matrix; b is the offset vector; f_a is the activation function; The subscripts t represents time.

The LSTM network is proposed to solve the problems of the gradient disappearance and gradient explosion, which owns long-term memory. The input gate g_t , output gate q_t, forget gate f_t and cell activation vector c_t are updated in the LSTM. The LSTM cell structure in a hidden layer is presented in Figure 3.

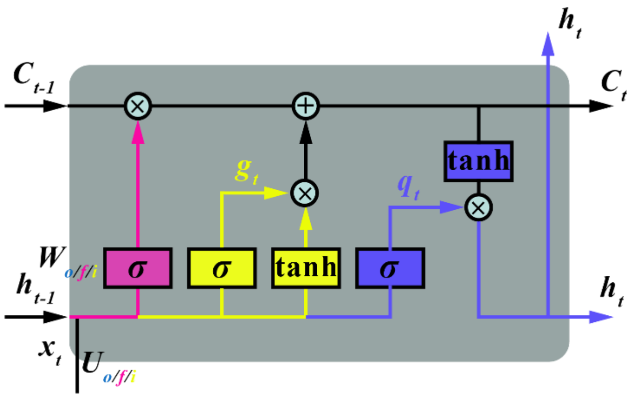


Figure 3: The LSTM structure

The updating equations are given as follows:

$$g_t = \sigma(\sum U_g x_t + \sum W_g h_{t-1} + b_g) \tag{13}$$

$$f_t = \sigma(\sum U_f x_t + \sum W_f h_{t-1} + b_f) \tag{14}$$

$$q_t = \sigma(\sum U_o x_t + \sum W_o h_{t-1} + b_o) \tag{15}$$

$$C_t = f_t C_{t-1} + g_t \tanh(\sum U_i x_t + \sum W_i h_{t-1} + b_i) \tag{16}$$

$$h_t = q_t \tanh(C_t) \tag{17}$$

where g_t, q_t, f_t and c_t are the input gate, output gate, forget gate and cell state respectively; W and b are the corresponding weight coefficient matrix and bias term, respectively; σ and tanh are the sigmoid and hyperbolic tangent activation functions, respectively.

2.4 Classification layer

The probability distributions of the representative features extracted by the 1-D CNN and BiLSTM layer, are fed into the fully connected layer for classification. Each output is mapped into a probability by a softmax function φ, which is defined by

$$\varphi(u_c) = e^{u_c} \Big/ \sum_{c=1}^{T} e^{u_c}, c = 1, 2, \dots, T \tag{18}$$

where $\varphi(u_c)$ is a T-dimensional probability vector and denotes the probability distribution under T kinds of test scenarios, u_c is the fusion features.

2.5 The proposed Multi-Scale CNNBiLSTM architecture

The proposed Multi-Scale CNNBiLSTM architecture consists of the multi-scale layer, the feature learning layer consisted of 1-D CNN, the BiLSTM layer and the classification layer. Figure 4 presents the proposed MSCNN-BiLSTM framework.

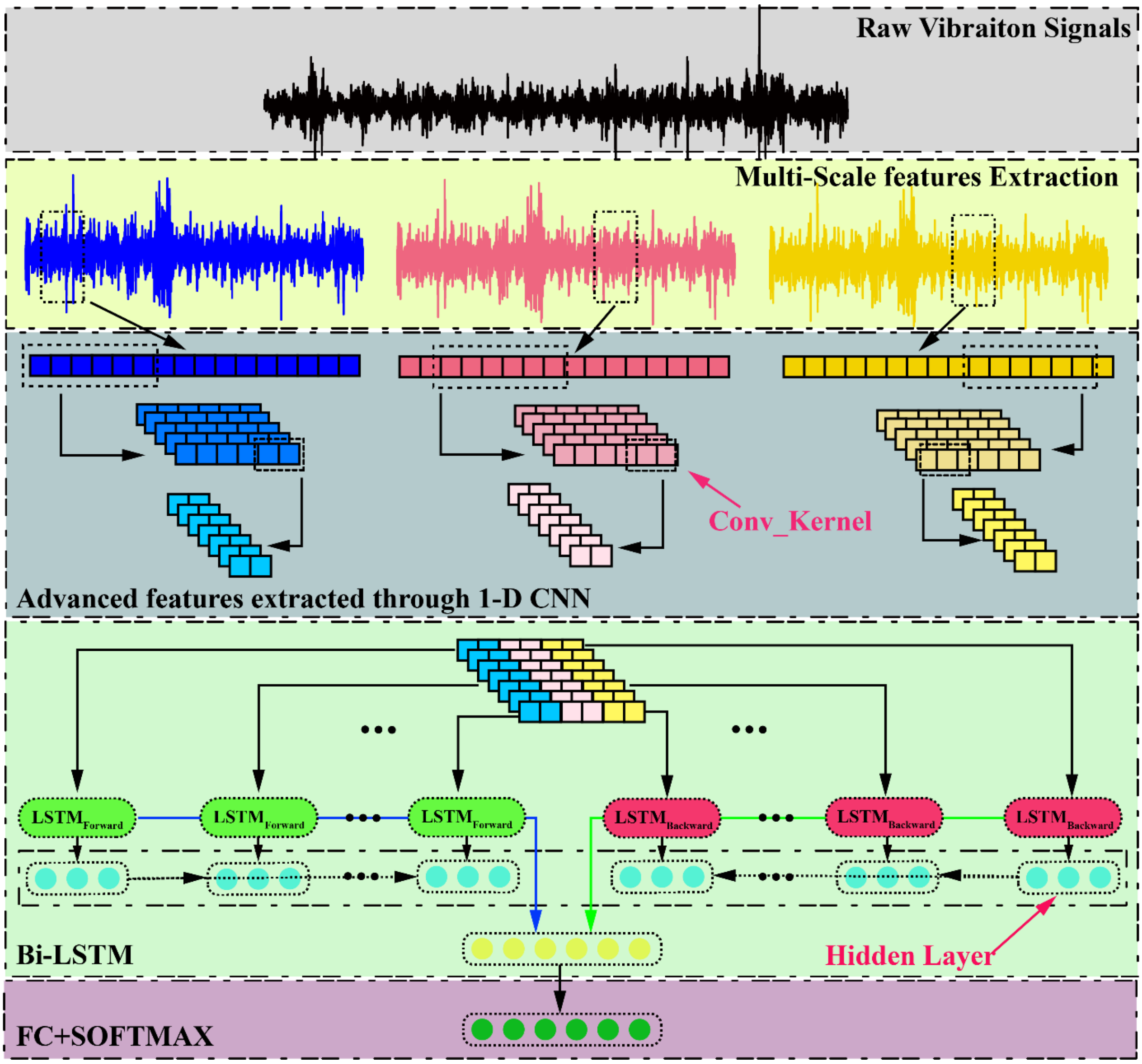


Figure 4: The architecture of the CMSCNN-LSTM model

As shown in Figure 4, the vibration signals measured by a sensor are fed into the MSCNN-BiLSTM network. The representation features with lower dimension features are obtained by the multi-scale layer and multiple parallels of 1-D CNN. The number of time steps of the advanced features fed into the BiLSTM network is decreased significantly from *n* to *L*, where *n* is the length of the input sequences, and *L* is the number of elements in the pooling layer. The relevant information is obtained by BiLSTM, which hidden in each advanced features of the forward and backward are fused into fully connected layer to calculate the probabilities of each working condition.

In contrast to the CNN-based model and LSTM-based model, the advantage offered by the structure of the proposed MSCNN-BiLSTM model is that its capability of examining time multi-scale features, which can capture more information needed to improve the performance of the model. The problem of high time complexity caused by being fully connected with the LSTM network has been improved by implementation of the pre-processing capability within using the advanced features as the feature vectors of a RNN network.

The improvement of the multi-scale coarse-grained procedure makes the data length of the sub-signals that obtained by the improved MS layer are the same as the original inputs. Which makes the feature extraction procedure using a CNN model to be more uniform and easier to be modified and maintained. The parameters of the feature extraction layer based on the 1D CNN in Figure 4 are presented in Table 2.

Table 2: The details of the 1D CNN in the CMSCNN-LSTM model

No.	Layers	Kernel Size/Stride	Filter numbers	Outputs Size
1	A sub-signal	-	-	[4096,1]
2	Conv_1	[128,1]/[5,1]	16	[794,1]
3	Pool_1	[64,1]/[3,1]	16	[395,1]
4	Conv_2	[2,1]/[2,1]	32	[111,1]
5	Pool_2	[3,1]/[2,1]	32	[54,1]
6	Conv_3	[2,1]/[1,1]	8	[18,1]
7	Pool_3	[3,1]/[2,1]	8	[7,1]

The parameters of the 1D CNN in the MSCNN-BiLSTM model are shown in details in Table 2. Compared with the CNN structure in other studies, the number of filters increases with the layers deepening. However, the number of filters of the last convolution layer is too small to reduce the time complexity. The BN layer and the ReLU activation are introduced between the convolution layer and the pooling layer to prevent the gradient disappearing.

The MSCNN-BiLSTM model is optimized by the Adam gradient descent optimization algorithm with a mini-batch size of 256 samples. The loss function is cross entropy. The learning rate is initialized to 0.001 with no decay on each update. A dropout layer is added before the fully connected layer to minimize over-fitting risk.

2.6 Weighted majority voting for multisensory diagnosis

It can be seen from review studies that the multisensory intelligent fault diagnosis is not only affected by the performance of a model [29], but how to summarize the useful information from multiple sensors also has significant impact on the final diagnostic results [30]. As a single model, the proposed MSCNN-BiLSTM has good performance. Therefore, in this study, motivated from the ensemble learning. Each MSCNN-BiLSTM model is regarded as a sub-model. Using different signals acquiring from different sensors as dataset to train MSCNN-BiLSTM model will integrate multiple learners to develop an improved deep learner to work in tandem. The weighted majority voting rule treats the predictions as the final class label. The choice of weights directly affects the final diagnostic result. Eq. (19) delineates weighted majority voting.

$$\mathrm{H}(x) = C_{arm\mathrm{ax}_j} \sum_{n=1}^{N} w_n h_n^j(x) \tag{19}$$

Where for each probabilities x, the prediction of N sub-model is $h_n^j(x)$. The weighted for majority voting of each $h_n^j(x)$ is w_n. The final prediction labels $\mathrm{H}(x)$ is calculated by the $C_{arm\mathrm{ax}_j}(\cdot)$ to find out which prediction has the most votes.

Weights play a key role in weighted majority voting rule. Thus, in this paper, GA [31] is used to find the optimized weights for majority voting. The motivation of using GA to evaluate weights is that we hope the weights can help to improve the F1 score of the ensemble MSCNN-BiLSTM framework. The fitness functions of the GA consist of F1 score. Assumed the MSCNN-BiLSTM model will solve a k-classification problem, the weights are $w_n^k = [w_n^1, w_n^2, \ldots, w_n^k]^{\mathrm{T}}$. The pseudo-code to solve the weight of the weighted majority vote is shown in Table 3

Table 3: Pseudo-code of weighted majority voting rule

Input：the output probabilities of each MSCNN-BiLSTM
Initial：initialization of the GA parameters，including N_{pop} , Proportion of cross variation (P_c, P_m) and the maximum iteration $N_{\max}$
Based on N_{pop} to Initialize the population and reset the number of iterations as $n=1$
while $n \le N_{\max}$ **perform**
Through the $w_n^k = [w_n^1, w_n^2, \ldots, w_n^k]^{\mathrm{T}}$ to calculate the weighted voting for each sensor
Calculate fitness through the F1 score
Choose according to competitive strategy
Perform crossover and mutation and renew the population
Determine convergence
if convergence **then**
jump out of the loop
end
end
end：Output the best weights for weighted majority voting

2.7 Fault diagnosis framework based on the MSCNN-BiLSTM for multisensory

In the section, the proposed MSCNN-BiLSTM model is examined using experiments on a bearing test rig. Figure 5 presents the fault diagnosis workflow based on the MSCNN-BiLSTM for multisensory.

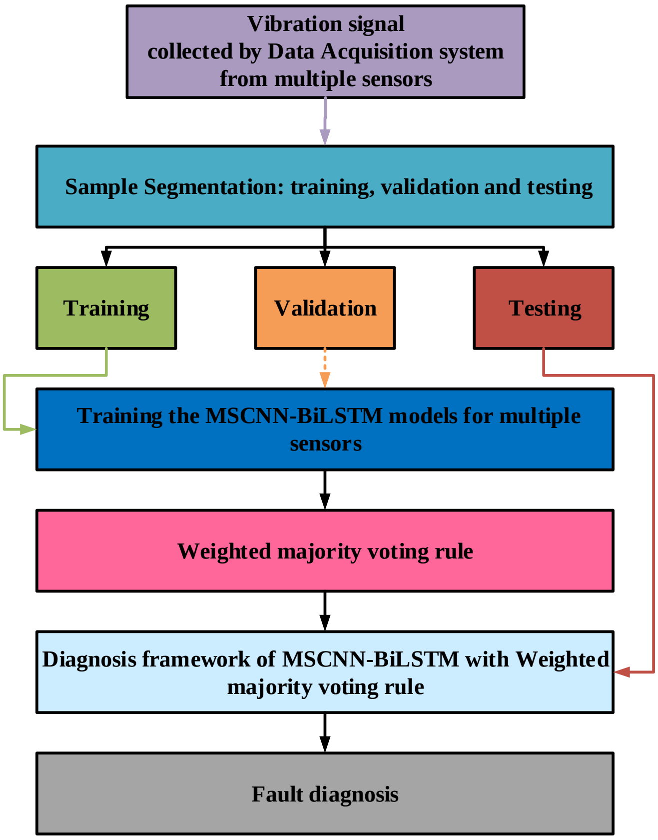


Figure 5: The fault diagnosis system based on the MSCNN-BiLSTM model with weighted majority voting rule

Figure 5 presents the intelligent diagnosis flowchart of wind turbine bearing based on MSCNN-BiLSTM model for multisensory. Using data acquisition system to obtain vibration signals from different sensors. Training dataset and validation dataset are built through sample segmentation and standardization. The best parameters of the MSCNN-BiLSTM model are trained and saved by cross validation. The number of the MSCNN-BiLSTM models are determined by the number of the sensors. For instance, use two sensors to collect signals can be used to train two MSCNN-BiLSTM models with different parameters. In order to integrate each MSCNN-BiLSTM model's performance, same as the ensemble learning, the predictions of the multiple MSCNN-BiLSTM models are fused in decision-making level through the proposed weighted majority voting rule for multisensory diagnosis.

3. Experiments and evaluation method

3.1 The description of experiment datasets

The bearing experimental data from Case Western Reserve University (CWRU) [32] and XJTU Xi'an Jiao Tong University (XJTU) [33] is used to construct different test scenarios to examine the performance of the proposed MSCNN-BiLSTM method. The NREL wind turbine transmission database is used to examine the practical application abilities in engineering of the proposed multisensory diagnosis method [34].

The CWRU experimental data, as the standard bearing vibration data set, is used to examine the performances of three kinds of RNN variants that include LSTM, BiLSTM and GRU and to compare the performance with CNN-based models for proving the superiority of the MSCNN-BiLSTM model. The data of CWRU covering normal state, inner race fault, ball fault and outer race fault in different azimuths (3, 6 and 12o'clock directions) are selected by two sensors with different sampling frequencies in order to validate the developed MSCNN-BiLSTM model. The data is examined for each of the fault category stated above. In total, 11 sets of data are used in this study. The motor loads range from 0 HP to 3 HP and the tested bearing model is SKF 6205.

The experimental data of CWRU is used to build different scenario and it is presented in Table 4.

Table 4: Datasets of bearing fault diagnosis for variable loads test

Datasets labels	Samples	Number of samples	Loads (hp)
I	Training	1600	0,1,2,3
	Test	160	0,1,2,3
II	Training	1600	0,1,2,3
	Test	160 (Added noise)	0,1,2,3

The data of XJTU covering inner race fault, cage fault outer race fault, and hybrid faults that consist of inner race, ball, cage and outer race failure, is selected by two sensors with different sampling directions to validate the developed MSCNN-BiLSTM model when the failures are weak. In the XJTU data includes the full-life of a bearing, the extrapolation abilities of the MSCNN-BiLSTM model are examination through the manually segmentation samples set that contain different damage magnitudes. Table 5 presents the details of the scenario setting for bearing degradation adaptation. The normal distribution of inner race fault from Phase 1 to Phase 3 are presented in Figure 6.

Table 5 Details of the bearing degradation configuration

Case Name	Training samples	Testing samples
A	Phase 1	Phase 2
B	Phase 1	Phase 3
C	Phase 2	Phase 1
D	Phase 2	Phase 3
E	Phase 3	Phase 1
F	Phase 3	Phase 2

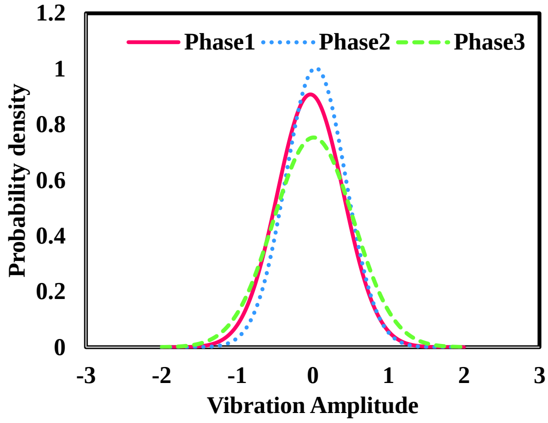
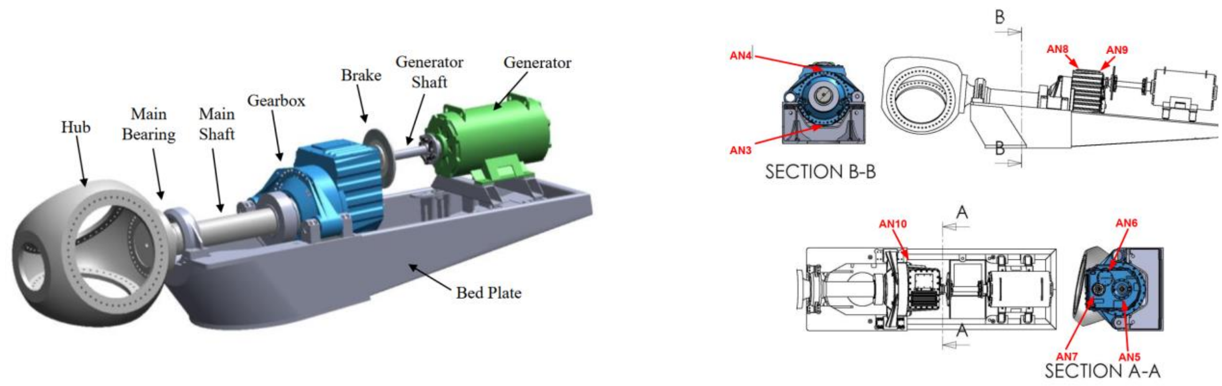


Figure 6: The normal distribution of different phases

As shown in Table 5 and Figure 6, six scenarios are constructed by extracting data from three different stages in the XJTU experimental process of bearing degradation. Phase 1, Phase 2 and Phase 3 respectively represent the development process of bearing failures from small to large, but it is worth noting that the data of the complete failure phase is not selected. The probability densities of the normal distributions of the three phases are different. Therefore, the bearing damage magnitude adaptation scenario test is used to verify the effectiveness of the proposed method.

Wind turbine condition monitoring benchmarking dataset provided by National Renewable Energy Laboratory is used to examine the proposed MSCNN-BiLSTM in the real engineering. The test turbine drive train configuration and the vibration sensor locations on wind turbine are shown in Figure 7 [35].



(a) Wind turbine drive train configuration (b) Vibration sensor locations

Figure 7 The illustration of the NREL test wind turbine

Table 6 presents a complied list of the actual damage occurred to the test drive train system. The damage detection are deemed through vibration analysis. In this study, HS-SH downwind bearing overheating, IMS-SH assembly damage of upwind and downwind bearings are used to build dataset for training MSCNN-BiLSTM model.

Table 6: Datasets of damage bearings of NREL wind turbine

labels	Samples	Number of samples	Sensors	Mode
1	Training	200	AN8	Healthy condition 1
	Test	200	AN9	
2	Training	200	AN5	Healthy condition 2
	Test	200	AN6	
3	Training	200	AN8	HS-SH downwind bearing overheating
	Test	200	AN9	
4	Training	200	AN5	IMS-SH downwind bearings damage
	Test	200	AN6	

3.2 Development environment and evaluation methodology

Different scenarios created based on the four aforementioned datasets are used to examine the proposed MSCNN-BiLSTM model. The data mining and setup of the deep learning model is conducted using the MATLAB® Deep Network Designer, MATLAB version 9.70 (R2019b, The MathWorks, Inc., Natick, MA, USA).

The F1 score is used to evaluate and compare the performance of the diagnosis model examined in this study, which offers a comprehensive metric to measure the extrapolation of the model. The definition of the F1 is presented in Eq. (20).

$$F1 = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{20}$$

where TP, FP, TN and FN mean correctly classified as positive samples, misclassified as positive samples, correctly classified as negative samples and misclassified as negative samples, respectively.

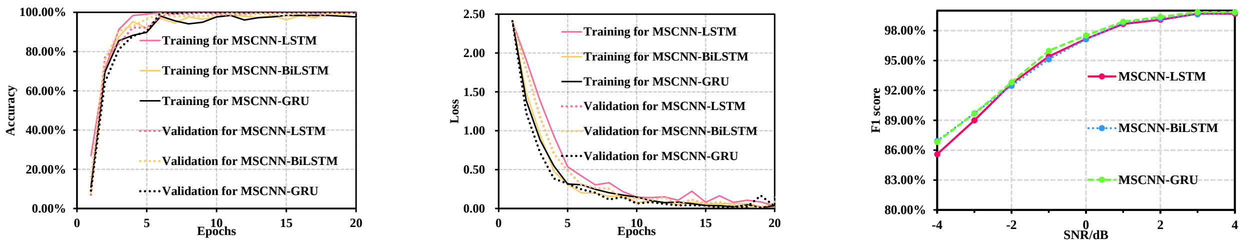
4. Validation and discussion

4.1 Comparison of the RNN variants

The LSTM module of the MSCNN-BiLSTM model is used to consider the long-term dependences of fault information. In this section, the influence of type of the RNN networks including the LSTM, BiLSTM [36] and GRU [37] is investigated using dataset II when integrated with the MSCNN model. The diagnosis models are named as MSCNN-LSTM and MSCNN-GRU. Figure 8(a) and Figure (b) show the accuracy and loss curves of training and validation of the three models. Figure 8 (c) presents the performance of the models examined using dataset II with SNR from -4dB to 4dB. Table 7 gives the training time, response time and F1 score of the methods examined in -4dB.

Table 7: Comparison of the performances of the MSCNN-LSTM model, the MSCNN-BiLSTM model and the MSCNN-GRU model

Methods	F1 score (%)	Training time (s)	Testing time (s)
MSCNN-LSTM	85.58 ±0.29	206.84+16.57	0.4943 ±0.0037
MSCNN-BiLSTM	86.93 ±0.17	209.85+19.45	0.4557 ±0.0035
MSCNN-GRU	86.83 ±0.21	204.42+16.93	0.4789 ±0.0038



(a) accuracy (b) loss (c) F1 score

Figure 8: The training and validation accuracy curves of the three models with increasing epochs and performances of the models examined using dataset II with SNR from -4dB to 4dB.

Figure 8 (b) presents the loss curves of the models' training and validation. The loss of the MSCNN-LSTM is the most unstable model among the three RNN-variants model, while the MSCNN-BiLSTM model is stable due to the fusion of forward and backward information. As shown in Table 4 and Figure 8, the differences between the training/response times of the variants are insignificant. The MSCNN-BiLSTM has a slightly longer training time and a shorter response time. That is because forward and backward propagation of advanced features to record different fault context information. It is noted that the MSCNN-GRU model examined in the 0 dB environment has the highest F1 score of 97.5%. This is because the GRU only contains update and reset gates. However, in a noisy environment (-4dB), the MSCNN-BiLSTM model has the highest F1 score of 86.9%, which is because the LSTM units can control whether the important information in them is retained or not. The two-way propagation makes the BiLSTM unit more capable of capturing long-term dependencies, which gives the model a better noise immunity. Thus, the MSCNN-BiLSTM model performs better in a large noise environment, which also proves the motivation of the proposed MSCNN-BiLSTM model.

4.2 Comparison with advanced methods

In order to confirm the superiority of the MSCNN-BiLSTM model in identifying the failure types and the failure magnitudes of the bearings in the noisy environments, Figure 10 provides the comparisons of the proposed MSCNN-BiLSTM, the LSTM [38], the CNN-LSTM [39], TICNN [40], MC-CNN [19], C-CNN [41], MA-CNN[42], MCCNN-LSTM[43], and MS-CNN [15] by using average F1 scores with 10 trails examined on dataset H with -4dB noise level to 4dB noise level. The diagnosis results of nine methods under different noise environments are shown in Figure 9 to further demonstrate the reliability of the proposed MSCNN-BiLSTM model. Table 8 presents the details of the F1 scores in average, nine methods examined by dataset II with 0 dB in 10 trails.

Table 8 Comparison of the Deep learning models

Methods	F1 score (%)
MSCNN-BiLSTM [Our method]	97.12±0.09
CNN-LSTM [2020]	92.03±0.24
MS-CNN [2019]	83.75±0.78

LSTM [2019]	83.08±1.49
TICNN [2018]	82.23±0.92
MC-CNN [2019]	81.19±1.92
C-CNN [2020]	75.83±1.14
MA-CNN [2020]	67.15±1.43
MCCNN-LSTM [2021]	96.00±0.15

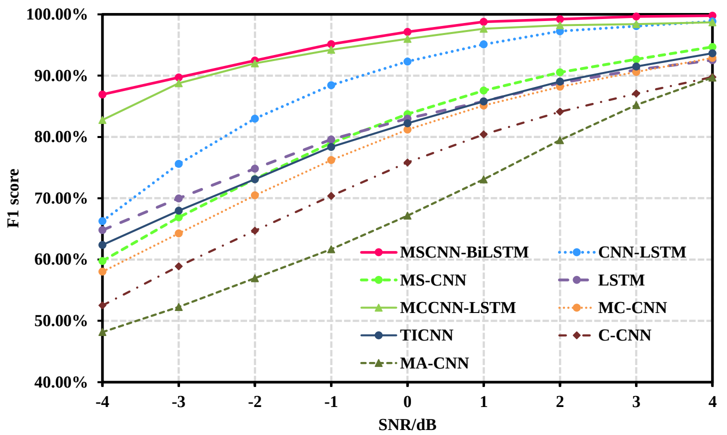


Figure 9: Diagnosis results examined in 10 trails on the noisy signals with different SNRs

As shown in Figure 9, the proposed MSCNN-BiLSTM model always shows the highest average F1 score with the SNR changed from -4dB to 4dB. The performance of the MCCNN-LSTM model is second only to the MSCNN-BiLSTM model. There are two possible reasons why the MSCNN-BiLSTM is better than the MCCNN-LSTM: 1. the multi-scale features extraction of the MSCNN-BiLSTM is developed based on multi-scale coarse-grained algorithm, which is more robust than the MCCNN-LSTM model using multi-scale convolution to extract multi-scale information. 2. The proposed MSCNN-BiLSTM considers the forward and backward fault semantics, which can capture more useful information than the MCCNN-LSTM only consider single direction fault semantics. Besides, The MA-CNN performs poorly when the test date source contains noise. That is because, the MA-CNN is used the gray images of the vibration signals as the inputs of the CNN-based model.

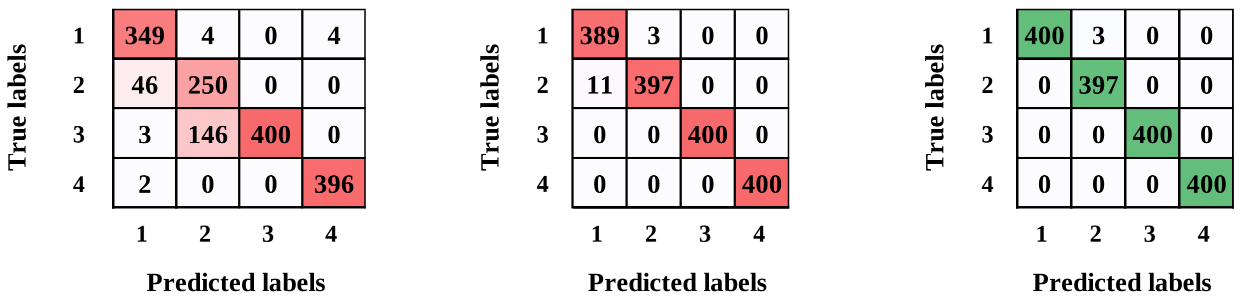
4.3 Damage magnitude detection adaptation scenario test for multisensory

In the damage magnitude detection adaptation scenario, the training dataset and the test dataset are from a same data source but diatribe differently due to the damage evolution, which can prove a strong extrapolation the proposed MSCNN-BiLSTM model has. Table 9 give the weights of majority voting and F1 score of each sensor.

Table 9: The weights of majority voting and F1 score for every conditions and sensor

	Weights for majority voting [Sensor1, Sensor2]	F1 Score (Sensor 1/ Sensor 2/ Fusion)
Inner Race fault	[1，1]	(0.9887/0.9962/**0.9963)**
Cage fault	[1，2]	(0.9197/0.9876/**1.0000)**
Outer race fault	[1，1]	(0.8503/1.0000/**1.0000)**
Hybrids faults	[1，2]	(0.9975/1.0000/**1.0000)**

As shown in Table 9, the examined results of MSCNN-BiLSTM model show that there are low F1 scores of the cage fault and outer rave fault for each sensor due to damage evolution and added noise (SNR=-4). Although the mean of F1 score of each sensor is high, there is a little false alarm rate that is needed to be avoided in fault diagnosis. The proposed weighted majority voting can integrate the positives of each sensor, through the weights votes, to improve F1 score of each classified conditions. It can be seen that the F1 score of cage fault for sensor 1 increases from 0.9197 to 1.0000, the F1 score of outer race fault for sensor 1 increases from 0.8503 to 1.0000. To further explain the mechanism of the weighted majority voting rule, confusion matrix is used to show how the weighted voting majority rule to improve the performance of the MSCNN-BiLSTM model for multisensory diagnosis. Figure 10 presents the diagnosis results, tested in different damage magnitude to examine the extrapolation of the MSCNN-BiLSTM model for multisensory.



(a) Sensor 1 (b) Sensor 2 (c) Fusion

Figure 10 Confusion matrix of diagnosis results

Figure 10 presents that the diagnosis result of sensor 1 has a very high false alarm rate for detecting cage fault, thus only 1 weight is given to sensor 1 for voting. In order to combine the excellent performance in distinguishing cage fault, 2 weights are given to sensor 2 for voting to reduce the false alarm. Therefore, there is little false alarm rate in the diagnosis result of the fusion.

The generalization of the proposed MSCNN-BiLSTM model for multisensory are examined through the damage evolution scenario, which is compared with the diagnosis results of multisensory information fused respectively on feature-level and majority voting decision-level. Figure 11

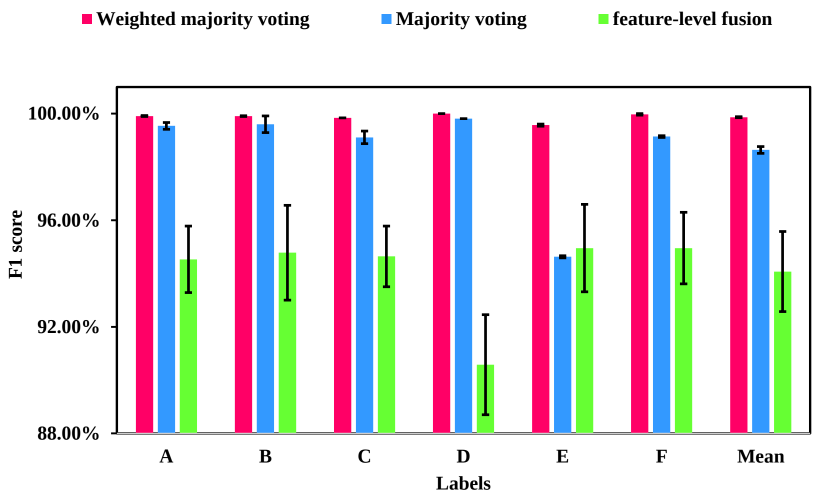


Figure 11: Comparison of the fusion methods for multisensory

As shown in Figure 11, the MSCNN-BiLSTM model fused the multisensory information on decision level through the proposed weighted majority voting rule has a better performance than the other methods when the examined under damage evolution scenario tests. The mean of F1 scores the weighted majority voting rule is higher than feature-level fusion for multisensory information by 5.7% and is higher than traditional majority voting by 1.23%. That indicates that the proposed weighted majority voting rule is effective.

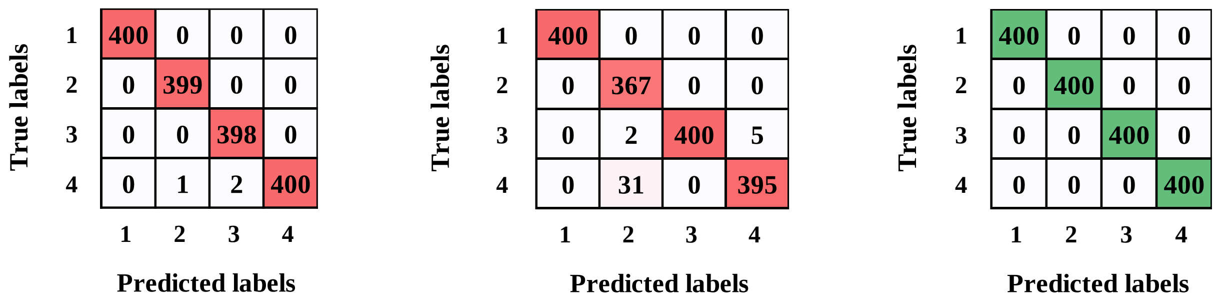
4.4 Real wind turbine fault diagnosis for multisensory

In real industrial engineering, the model for multisensory diagnosis is offline trained in advanced. Thus, the data of NREL wind turbine respectively acquired on different days. In the NREL dataset, the data acquired on day 1 is used to train the MSCNN-BiLSTM model, and test the model respectively through day 2 and day 3. Confusion matrix and the F1 score are used to demonstrate the superiority of the MSCNN-BiLSTM and to explain the mechanism of the weights majority voting. Table 10 give the weights of majority voting and the F1 score of each sensor.

Table 10: The weights of majority voting and F1 score for every conditions and sensor

	Weights for majority voting [Sensor1, Sensor2]	F1 Score (Sensor 1/ Sensor 2/ Fusion)
Healthy condition 1	[1，2]	1.000/1.000/**1.000**
Healthy condition 2	[1，2]	1.000/1.000/**1.000**
downwind bearing overheating	[1，2]	1.000/0.9913/**1.000**
downwind bearings damage	[1，1]	0.9963/0.9622/**1.000**

As shown in Table 10, there are some false alarm rate in the case of only using a sensor to diagnosis. The faults of downwind bearing overheating and damage sometime are misclassified due to a sensor can not capture useful information for fault diagnosis. The proposed weighted majority voting rule can integrate the diagnosis results from multiple sensors for improving the performance of the diagnosis. The sensor 2 takes up a weight 2 votes for healthy condition1, healthy condition2 and bearing overheating, which can deal with some false alarm rate existed between the condition of bearing overheating and damage. To further demonstrate the performance of the proposed weighted majority voting, the diagnosis results of the single sensor and multisensory are respectively presented by confusion matrix in Figure 11.



(a) Sensor 1 (b) Sensor 2 (c) Fusion

Figure 12 Confusion matrix of diagnosis results

As shown in Figure 12, the proposed weighted majority voting rule reduce the false alarm rate existing between the healthy condition 2 and bearing damage. The mechanism of the weighted majority voting rule is that Sensor 2 is given more vote weight to health condition 1, health condition 2 and bearing overheating fault to compensate for the bias of diagnosis results on bearing damage. It can be seen that although there are false positives in both sensor 1 and sensor 2 during diagnosis, these false alarms are eliminated after weight voting, which proves the reliability of the proposed method in real wind turbine engineering.

5. Conclusions

A novel fault diagnosis method of wind turbine bearings is developed based on multi-scale coarse-grained procedure algorithm, CNN, BiLSTM and a proposed weighted majority voting rule for multisensory fault diagnosis. The method is combined with the advantages of the CNN in auto features extraction and BiLSTM in capturing the correlation features. CNN is used to extract useful advanced fearless from the multi-scale sub-signals that generated by an improved multi-scale coarse-grained procedure algorithm, which also can reduce the dimension of the fault features to decrease the calculated amounts of the LSTM unit. In addition, a weighted majority voting rule is designed to fuse the multisensory information in the decision-fusion, which improves the robustness of the MSCNN-BiLSTM model. The verification of our method are examined through multiple groups of experimental data and the main conclusions of this study are as follows:

(1) The robustness of the MSCNN-BiLSTM model can be improved by using bidirectional LSTM network to capture forward and backward semantic information between advanced fault features, which has higher diagnosis performance than MSCNN-GRU and MSCNN-LSTM when they are examined in noisy environment.

(2) Compared with existing fault diagnosis model developed based on CNN network, the proposed MSCNN-BiLSTM model has the highest F1 score by 97.12% examined through anti-noise test. The generalization of the proposed MSCNN-BiLSTM model is better than the generalizations of the LSTM,

the CNN-LSTM, TICNN, MC-CNN, C-CNN, MA-CNN, MCCNN-LSTM and MS-CNN.

(3) The proposed weighted majority voting rule can take advantages of a good fault diagnosis results of each sensor to improve the final diagnosis performance. In the damage evolution test scenario, the 0.8503 F1 score of the outer race fault, diagnosed by a sensor, is improved by the proposed weighted majority voting rule to increase to 1.0000, which is helped by giving another sensor more vote weights.

(4) The proposed weighted majority voting rule is compared with different methods for multisensory diagnosis, that include a traditional majority voting rule that belongs to fusion on decision-level and fusion on feature-level. The results indicate that the proposed weighted majority voting rule is higher than the others by 1.23% and 5.7%.

Acknowledgements

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (grant numbers: 51676131, 51875361 and 51976131). Shanghai Pujiang Program (2019PJD054). Science and Technology Commission of Shanghai Municipality (grant number: 1906052200), Royal Society (grant number: IEC\NSFC\170054), the National Renewable Energy Laboratory and the United States Department of Energy for providing the benchmarking datasets of wind turbine gearbox vibration condition monitoring.

References

[1] Mbungu, Nsilulu T , et al. "Optimisation of grid connected hybrid photovoltaic–wind–battery system using model predictive control design." Iet Renewable Power Generation 11.14(2017):1760-1768.
[2] Chen, J. , et al. "Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals - ScienceDirect." Renewable Energy 89(2016):80-92.
[3] Teng, W. , et al. "Compound faults diagnosis and analysis for a wind turbine gearbox via a novel vibration model and empirical wavelet transform." Renewable Energy 136.JUN.(2019):393-402.
[4] Xu, Z. , C. Li , and Y. Yang . "Fault diagnosis of rolling bearing of wind turbines based on the

Variational Mode Decomposition and Deep Convolutional Neural Networks." Applied Soft Computing 95(2020):106515.
[5] Xu, Z. , C. Li , and Y. Yang . "Fault diagnosis of rolling bearings using an Improved Multi-Scale Convolutional Neural Network with Feature Attention mechanism." ISA Transactions (2020).
[6] Li, B. , et al. "Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review." Renewable and Sustainable Energy Reviews 138(2020):110512.
[7] Jing, L. , et al. "An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox." Sensors (Switzerland) 17.2(2017):414.
[8] Cao, L. , et al. "Fault Diagnosis of Wind Turbine Gearbox Based on Deep Bi-Directional Long Short-Term Memory Under Time-Varying Non-Stationary Operating Conditions." IEEE Access 7(2019):1-1.
[9] Wang, Z. , J. Wang , and Y. Wang . "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition." Neurocomputing 310.OCT.8(2018):213-222.
[10] Wang, Tianyang , et al. "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review." Mechanical Systems and Signal Processing 126.JUL.1(2019):662-685.
[11] H Zuo, et al. "Effects of different poses and wind speeds on wind-induced vibration haracteristics of a dish solar concentrator system." Renewable Energy 168(2021).
[12] Huang, Y. , M. Gu , and M. Naggar . "Effect of soil-structure interaction on wind-induced responses of supertall buildings with large pile groups." Engineering Structures 243.2(2021):112557.
[13] Wang, J. , et al. "A multi-scale convolution neural network for featureless fault diagnosis." 2016 International Symposium on Flexible Automation (ISFA) IEEE, 2016.
[14] Chen, Z. , K. Gryllias , and W. Li . "Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine." Mechanical Systems and Signal Processing 133(2019).
[15] Jiang, et al. "Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox." IEEE Transactions on Industrial Electronics (2019).
[16] Zhao, B. , et al. "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions." Knowledge-Based Systems 199(2020):105971.
[17] X Wang, D. Mao , and X Li. "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network." Measurement 173.6(2021):108518.
[18] Zhang, W., et al. "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load." Mechanical systems and signal processing(2018):439-453.
[19] Huang, W. , et al. "An improved deep convolutional neural network with multi-scale information for bearing fault diagnosis." Neurocomputing 359(2019):77-92.
[20] Zhao R. , Yan, J R, Wang, K. Mao, Learning to monitor machine health with convolutional bi-directional lstm networks, Sensors 17 (2) (2017) 273.
[21] Lu, W. , et al. "Early Fault Detection Approach With Deep Architectures." IEEE Transactions on Instrumentation and Measurement (2018):1-11.
[22] Jing, L. , et al. "An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox." Sensors (Switzerland) 17.2(2017):414.
[23] Ma, A , et al. "Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis." Mechanical Systems and Signal Processing 144.
[24] Chen, Z. , and W. Li . "Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network." IEEE Transactions on Instrumentation and Measurement (2017):1-10.
[25] Liu, J. , et al. "An integrated multi-sensor fusion-based deep feature learning approach for rotating

machinery diagnosis." Measurement Science & Technology (2018).
[26]Shao, H. , et al. "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance." Information Fusion (2021).
[27]Hochreiter, S. , & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9 (8), 1735-1780.
[28]Lin, Z. , et al. "Coordinated pitch & torque control of large-scale wind turbine based on Pareto efficiency analysis." Energy 147.MAR.15(2018):812-825.
[29]Wang, Z. , H. Huang , and Y. Wang . "Fault diagnosis of planetary gearbox using multi-criteria feature selection and heterogeneous ensemble learning classification." Measurement 173.5(2020):108654.
[30]Huang, R. , et al. "Deep Ensemble Capsule Network for Intelligent Compound Fault Diagnosis Using Multisensory Data." IEEE Transactions on Instrumentation and Measurement 69.5(2020):2304-2314.
[31]Zwickl, D. J. . "Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion." Dissertations & Theses - Gradworks 3.5(2008):257-260.
[32]Wade, et al. "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study." Mechanical Systems and Signal Processing 64-65(2015):100-131.
[33]Wang, B. , et al. "A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings." IEEE Transactions on Reliability (2018):1-12.
[34]Sheng, S. . "Report on Wind Turbine Subsystem Reliability - A Survey of Various Databases (Presentation)." Office of Scientific & Technical Information Technical Reports (2013).
[35]Sheng. "Investigation of Various Condition Monitoring Techniques Based on a Damaged Wind Turbine Gearbox." (2011).
[36]Ma, X. , and E. Hovy . "End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF." (2016).
[37]Gal, Y. , & Ghahramani, Z. . (2015). A theoretically grounded application of dropout in recurrent neural networks. Stats, 285-290.
[38]Lei, J. , Liu, C. , & Jiang, D. . (2019). Fault diagnosis of wind turbine based on long short-term memory networks. Renewable Energy, 133(APR.), 422-432.
[39]Hao, S. , et al. "Multisensor Bearing Fault Diagnosis Based on One-dimensional Convolutional Long Short-Term Memory Networks." Measurement 159(2020):107802.
[40]Zhang,W., et al. "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. " Mech Syst Signal Process (2018):439–453.
[41]Chang, Y. , et al. "Intelligent fault diagnosis of Wind Turbines via a Deep Learning Network Using Parallel Convolution Layers with Multi-Scale Kernels." Renewable Energy 153(2020).
[42]Wang, H. , et al. "Intelligent Bearing Fault Diagnosis Using Multi-Head Attention-Based CNN." Procedia Manufacturing 49(2020):112-118.
[43]Chen, X. et al. "Bearing fault diagnosis base on multi-scale CNN and LSTM model." Journal of Intelligent Manufacturing 32(2021).