



An intelligent of fault diagnosis and predicting remaining useful life of rolling bearings based on convolutional neural network with bidirectional LSTM

KUMARAN BHARATHEEDASAN^{1,*}, TANMOY MAITY¹, L A KUMARASWAMIDHAS¹ and MURUGANANDAM DURAIRAJ²

¹Department of Mining and Machinery Engineering, Indian Institute of Technology (ISM), Dhanbad 826004, India

²Department of Mechanical Engineering, Adhi Engineering College, Chennai 631605, India
e-mail: bharatheedasan.17dp000279@mme.iitism.ac.in; tanmoy1826@gmail.com; lakdhas1978@gmail.com; murudurai@gmail.com

MS received 19 October 2022; revised 19 October 2022; accepted 15 April 2023

Abstract. The importance of the quality of life of rotating machinery increases the Bearing fault diagnosis. Deep learning models (DL)-based databases become increasingly smart in the field of fault diagnostics, the latest research has widely used CNNs (convolutional neural networks). This paper proposes a new way to diagnose bearing failures with CNN with Bilinear LSTM. Traditional CNNs are however not easy to detect defects due to the fixed geometry of complex fault diagnosis with different working conditions. Our primary and secondary classifiers at specified layers replace primitive shape convolutions with reconfigurable convolutions, resulting in classification results with stringent feature time-frequency incompatibility and a larger receptive field. To acquire more adaptive knowledge and insight into the proposed approach, we employ the CWRU (Case Western Reserve University) opensource dataset to compare classification accuracy. The bearing dataset has been subjected to comprehensive experiments and evaluations in order to confirm the efficacy of the suggested technique's diagnostic performance in a variety of settings. By comparing multiple perspectives on the same dataset with related tasks, the proposed method's superiority is proved. To limit the effect of noise and avoid temporal oscillations, degraded index sequences are matched with a CNN. Current and previous inspection data are fed into a new CNN-BiLSTM model, which is then used to predict the useful time and compatible power values of bearing RULs. When it comes to output, go with the lifetime percentage. The proposed method has been tested by accelerating bearing operation to failure, and the results show that the method has advantages in predicting RUL more accurately. The results of the experiments suggest that the proposed core distance measurement method is a viable new tool for intelligent rolling bearing diagnosis. The BiLSTM technique is more diagnostic than some generic models, according to experimental results using the 48 K and 12 K CWRU datasets, with overall accuracy of 99.80% and 98.3%, respectively.

Keywords. Deep learning models (DL); CNNs (convolutional neural networks); CWRU (Case Western Reserve University); CNN-BiLSTM model; RUL.

1. Introduction

Bearings play a crucial role in rotating machinery. One of the most common causes of motor failure is bearing problems, and discovering failures early on can save time and money in the long run [1]. Many scientific domains have greatly boosted their use of ML or DL in recent years. One of the most fascinating and widely used fields in the real world is intelligent defect detection.

Developing a network design that can give good diagnostic performance in a reasonably short amount of time is a fundamental issue in applying machine learning approaches to bearing fault diagnosis [2]. Signal processing techniques are used to accomplish data-driven intelligent bearing defect detection first. These are “vibration signals” or “motor current signals,” respectively, as recorded by an accelerometer or a frequency converter [3]. Vibration signals have been shown in the literature to be more useful for obtaining more precise findings [4].

Features must be extracted and used from learning algorithms developed to achieve maximum accuracy to use

*For correspondence
Published online: 08 July 2023

ML or DL techniques for bearing failure detection. The function has three domains: time [5], frequency [6], and time [7].

The focus of this publication is on rolling bearing fault diagnostics. The performance, stability, and life cycle of rotating machinery are all affected by bearing condition assessments. Because the working environment of rolling bearings is typically sealed and variable, complicated and effective defect diagnosis procedures are difficult to develop [8]. To check the state of rolling bearings that are inaccessible without disassembly, however, external relevant information must be analysed. Because the vibration signal's immediate value is continually changing, the rolling bearing signal's properties alter with time. As a result, developing a fault diagnosis system with a self-learning feature that can learn signal properties on a continual basis is critical [9].

Rolling-element bearings are a common rotary machinery component, and rolling bearings defects affect the operation of normal rotary machinery which causes serious equipment damage, economic costs, and sometimes life-threatening damage. For mechanical and intelligent integration systems, safety and reliability have become increasingly important in light of the rapid development of modern industries [10]. Reliability of a rotating machine is directly related to the performance of its main supporting component, the roller bearing. As a result, developing an intelligent diagnostic approach to discover and diagnose faults as soon as feasible is critical [11].

1.1 Contribution of the paper

1. This work shows how to use the CNN-BiLSTM approach to diagnose motor bearing faults without the requirement for sophisticated data pretreatment or signal processing.
2. Signal feature extraction is needed by the fault diagnostic method. This research proposes a convolutional neural network-based fault identification approach. It is a full-featured diagnostic procedure that eliminates the need for manual feature extraction and problem detection.
3. Explore the model's unique process for learning and classifying mechanical functions by using CNN-BiLSTM to show the learned function.
4. The results of the experiments show that the method can extract appropriate fault features from raw bearing vibration data and classify defects with good accuracy and stability.
5. A CNN-BiLSTM neural network and adaptively learnt deep features are used to develop a unique RUL prediction algorithm in two healthy states. This method takes advantage of not only the bearing faults' good representation ability, but also the deterioration process' temporal information.

This paper is structured as follows. The literature survey is in section 2. The network architecture employed in this study, CNN-BiLSTM, is briefly discussed in section 3. Section 4 describes experiments, data sets, defect classification and hyperparameter selection of the proposed CNN-BiLSTM architecture. Section 5 discusses and compares our method to other methods and section 6 finally provides the conclusions and future work.

2. Literature survey

Various “shallow” machine learning and data mining algorithms, as well as artificial neural networks (ANNs), were utilised more recently, prior to the deep learning boom. Using these algorithms necessitates a high level of knowledge as well as intricate feature engineering [12]. Principal component analysis (PCA) for obtaining the next feature from the first dataset is an example of detailed data analysis that leads to dimensionality reduction. Last but not least, the ML algorithm is a common feature.

One of the very first analyses could be found on the use of artificial intelligence (AI) technology to diagnose engine failures [13]. This is a systematic summary of the error frequencies of different types of motor defects and associated ANN-using papers.

In the last 30 years, ANN has been used to diagnose problems as one of the first artificial intelligence modelling techniques [14]. The motor bearings are reflected in the damping coefficient B based on the nonlinear stator current map “ I ” and the rotor speed “ R .” By creating a mapping neural network using the stator current and motor speed as inputs and the anticipated bearing as output, you can prevent this nonlinear mapping. Under various operating conditions, the Dayton 6K bearing lab test rig recorded 35 training models and 70 test data models. An existing neural network with two input nodes achieves 94.7% maximum accuracy in the carrying defect detection [15].

One of the initial PCA adoptions for the diagnosis of failure can be found in [16]. The experiment demonstrated that it is significant to improve the accuracy of a failure diagnosis from 88% to 98% with the use of only the PCA identification feature instead of 13 unique features. During this study, we have clearly shown that the PCA method is the most accurate and least efficient of all features in defect classification. Other PCA works [17–19] also use data mining to expedite manual classification and create more representative features.

The k-NN algorithm output is a class of objects in the k-NN classification, identified by the majority votes of k 's closest neighbors. An early implementation of the k-NN classification is available in [20]. In this respect, k-NN emerges as an essential algorithm for identifying ceramic bearing defects for audio-based data mining. In addition,

k-NN [21] other research studies use k-NN to evaluate each new data sample from a certain class of defects.

SVM is a model map learning which analyses data used for the analysis of non-stochastic grading or regression. The classic work you did to identify room defects by using SVMs is shown in [22]. The classification results generated by SVM are generally better than ANN's performance and in all cases are optimal. Similar papers from SVM also described the efficiency and effectiveness of the use of SVMs by defect categorizers.

Many auto-encoder studies that apply to failure-diagnosis [23] can find early attempts to make an adaptive failure of five-layer auto-encoder-based DNN. Frequency Spectrum Characteristics understanding and effective classification for backpropagation neural networks, classification accuracy is 99.6%.

For the detection of faults in inner and outer rings, as well as rotating elements of aviation engine shafts, vibration measurements are included in the dataset [24]. A Gaussian radial kernel function is used to create the new AE. Stacking AEs will be created with this new AE and certain existing AEs. The average accuracy is 86.75%, which is significantly better than SAE (44.90%) and DBN (19.65%).

A 4-layer DNN has been implemented because of the large data volume required for training deep neural networks with a 70% compression stacked sparse autoencoder [25], on the basis of our proposed training. Only 30% of the data shows it is necessary. In order to compress vibration data in the deformed space and extract adaptive functions from it, the nonlinear projection is carried out, and the proposed method's accuracy is 97.47%, 8% higher than SVM and 60% higher than 3-layer ANN. Multilayer artificial neural networks are 46 percent more expensive.

In the output layer, a clustering technique is used to determine the likelihood that a sample belongs to Fusel Maximum Likelihood Optimization (FMLE) [26], in addition to the most common Softmax classifier. The SDAE 8-layer still extracts useful features in vibration signals, while GG identifies different types of defects. The classification accuracy in the worst case is 93.3%, nearly 10% higher than the current EMD-based method of extraction.

A DL-based model for error diagnostics by using an autoencoder [27]. A smooth voting approach that adds the predictions for each signal segment clipped by a sliding window has been devised to increase classification accuracy and stability. Adding white Gaussian noise to the original CWRU dataset to generate a custom dataset for diagnosing performance in a bright environment [28].

Two-dimensional image recognition and diagnosis have been revolutionised by convolutional neural networks [6]. The synthetic product neural network is a "end-to-end" network structure that uses a neural network to finish the entire process of feature extraction, feature dimensionality reduction, and classifier classification. These properties of convolutional neural networks can undoubtedly

compensate for the deficiencies of present defect detection methods, provide new ideas for existing defect detection algorithms, and improve the diagnostic performance of existing defect detection algorithms [29].

In practise, intelligent algorithms-based defect diagnosis methods may successfully overcome the limitations of signal analysis methods and have a higher recognition rate than traditional approaches. In rolling bearing fault detection, however, the most intelligent algorithmic methods still exist. This is because to the network model's depth, which allows it to extract information accurately enough for diagnostic purposes. The use of temporal data on defect features is predicted to increase the accuracy of bearing RUL predictions over time as bearings deteriorate.

2.1 Objective

The focus of this study is on the intelligent detection of rotating electrical machine defects, with reference to deep learning research development. The advantages of the proposed technique include:

1. Real-time bearing condition detection using a dense network structure and raw data input.
2. Learn functions automatically from raw signals without any preprocessing (EMD, HHT, etc.).
3. Utilizes minimal datasets to provide effective training and classification methods.
4. Cut maintenance expenses by eliminating unnecessary interventions and inspections.

Experiments show that the suggested model outperforms various modern intelligent fault identification methods in noisy situations.

2.2 Rolling bearing fault signature

A typical roller bearing construction consists of an outer ring and an inner ring mounted on a bearing housing. It can be attached to rotating shafts, rolling elements and support cages. Roller bearings are the most fragile parts of rotating machines. If damaged, the wear of the internal balls can cause the bearing to drop dramatically. Various root causes of bearing failure, such as improper sizing, leakage and excessive loads. When a bearing is defective, it causes periodic abnormal vibrations, the magnitude of which depends on the type of defect. The magnitude of the defect, the shaft speed, the load, and the position of the defect all influence the failure vibration frequency characteristics, as do several common vibration frequencies for different mechanical defect features, such as:

$$f_c = \frac{f_r}{2} \left(1 - \left(\frac{d}{D_m} \cos \alpha \right) \right) = \frac{f_{bi}}{N_b}$$

$$\begin{aligned}
 f_{be} &= \frac{f_r}{2} N_b \left(1 - \left(\frac{d}{D_m} \cos \alpha\right)\right) \\
 f_{bi} &= \frac{f_r}{2} \left(1 + \left(\frac{d}{D_m} \cos \alpha\right)\right) \\
 f_b &= \frac{f_r D_m}{2d} \left(1 - \left(\frac{d}{D_m} \cos \alpha\right)\right)^2
 \end{aligned} \quad (1)$$

where D_m is the pitch diameter; d is the roller diameter; N_b is the number of rollers; α is the ball contact angle; and f_c , f_{bi} , f_{be} , and f_b are the cage, inner race, outer race and roller fault frequencies, respectively.

3. Convolutional neural network (CNN)

A CNN is made up of many convolutional layers, normalisation (BN) layers, activity layers, pooling layers, and classification layers as a multi-level neural network. Pooling layers create input and output features whereas convolutional layers mix the data with a collection of unknown filters known as kernels. During training, each kernel is a vertical matrix of dimensions in which the elements (or hyperparameters) are taught. To extract local input features, the kernel is employed. Because the same kernels are used to alter the input units of each layer, CNNs tend to have significantly smaller hyperparameters during training than ANNs, reducing the risk of consensus overload.

The study uses one-dimensional convolution since the vibration signal is a one-dimensional time series.

$$y_i^{l+1} = K_i^l \times x^l(j) + b_i^l \quad (2)$$

where k and b represent the weights and biases of the i^{th} filter and the l^{th} layer, respectively, and x represents the j^{th} local input of the l^{th} layer.

The collected features are selected and filtered using a pooling layer after hierarchical synthesis. To get local properties of the region and limit the number of parameters, max pooling chooses the maximum statistics for the most common pooling processes. Following is an interpretation of the $(l+1)$ th layer of the i th channel after pooling.

$$P_i^{l+1} = \max_{(j-1)W+1 \leq t \leq jW} \{q_i^l(t)\} \quad (3)$$

where q represents the t^{th} neuron in the l^{th} layer of the i^{th} channel and w is the width of the pooling kernel.

As shown in figure 1, the proposed CNN approach comprises two main steps. The first step is to filter the raw input signal into the frequency domain. The second phase feeds the input signal to the CNN. The following are explained in more detail in these two steps. The first step is to construct a frequency/frequency resolution level based on a quick curtogram analysis. The highest kurtosis signal is then chosen as the filtered signal. This is essential to the

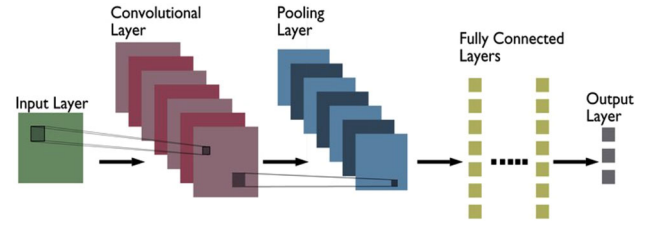


Figure 1. CNN structure.

proposed error diagnostic approach, as the raw input signal has a signal that is generally uncontrollable at a high noise ratio. This signal preprocessing can minimize the noise components of the input signal in the proposed CNN scheme. The filtered signal is then demodulated to remove carrier frequencies independent of its characteristic frequency. Time serial signals pre-processed before the CNN multichannel model is input are converted into the frequency domain. The CNN considers multiple sensor data channels as inputs for the multi-channel CNN model to be essential. These various sensor signals are a time series that is not fully synchronized (that is, not collected at exactly the same time). This means that there is an unknown delay between measuring samples on one channel and samples on other channels (in milliseconds). Because of the unique asynchronous structure of time series data, feeding raw time series data directly to a CNN risks losing information.

3.1 Long short term memory (LSTM)

In RNN trainings where input and output gate functions are introduced, LSTM forgetting doors can avoid gradient explosions and gradient disappearances. The RNN training problem can be solved with three gate functions. An LSTM cell is shown in figure 2.

The following is an example of an LSTM different cells equation:

$$i_t = \sigma(W_{x_i} x_t + W_{h_i} h_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{x_f} x_t + W_{h_f} h_{t-1} + b_f) \quad (5)$$

$$o_t = \sigma(W_{x_o} x_t + W_{h_o} h_{t-1} + b_o) \quad (6)$$

$$c_t = \tanh(W_{x_c} x_t + W_{h_c} h_{t-1} + b_c) \quad (7)$$

$$c_t = f_t e_{t-1} + i_t e_{c_t} \quad (8)$$

$$h_t = o_t \tanh(c_t) \quad (9)$$

X_t indicates the value of the t^{th} time series given to the LSTM. c_t represents the LSTM core memory cell. The storage unit can be used to regulate the conversion of various types of temporal data. The gate is responsible for determining what information will be presented next. The memory gate displays the amount of data that has been

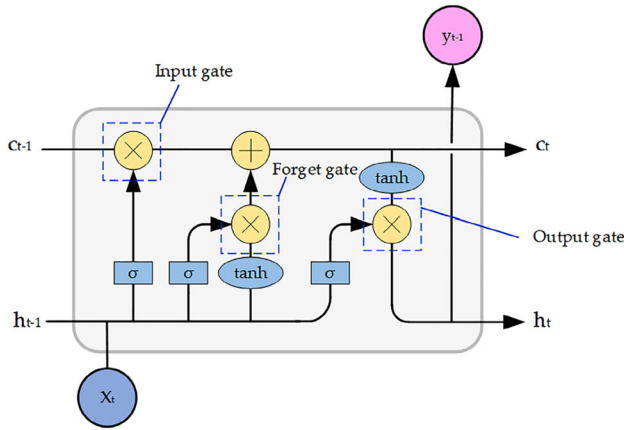


Figure 2. LSTM model.

retained since the last time you used it. The process output from the current state to the next state is determined by the output gate.

Finally, the softmax function transforms the neuron output into a probability distribution for each of the 10 roller bearing states. Where z denotes the output of the j th neuron, the softmax function is expressed as follows:

$$q(z_j) = \text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{10} e^{z_k}} \quad (10)$$

3.2 Bidirectional LSTM (BiLSTM)

BiLSTM (Bidirectional Long Short Term Memory) is a forward and reverse LSTM that can link predictions based on both forward and reverse data. Only time-related data can match LSTMs in one direction. BiLSTM has added an inverse LSTM to capture LSTM drop patterns for BiLSTM. Figure 3 shows the structure of BiLSTM. L_i is the LSTM forward, L'_i is the LSTM opposite, s is the LSTM time series.

The vibration signal data is fed into the network using CNN, which uses its convolutional layer and grouping layer

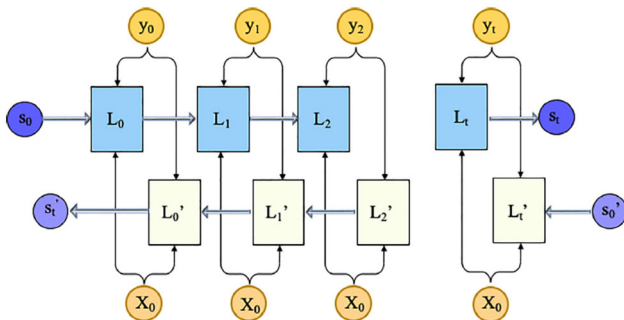


Figure 3. Bi-LSTM structure.

to minimise data dimension and analyse fault features. The attention hierarchy computes feature weighting factors in an adaptable manner. The features that are kept to be learned in the CNN are listed in the first row, and the adaptive weight matrix is found after Softmax in the second row. Divide the two rows' results by two. To put it another way, to get the weight attribute matrix, you must first weight the features. Effective features generally affect classification results significantly, while ineffective or inefficient characteristics have less impact on classification results.

Weighing characteristics can improve efficiency and decrease ineffective or inefficient characteristics. Weighted Hierarchical Networks (WHNs) and CNN Attention to learn more about feature temporal correlation, hierarchical features are delivered to LSTM networks. The training results are then sent over a fully connected network to the Softmax layer in order to finally get the rating results. With regard to defect characteristics as well as time features of the original signal, the BiLSTM model emphasizes how important it is to consider each attribute separately. This method can increase the accuracy of failure diagnosis by learning from high-dimensional raw data more efficiently.

Figure 4 depicts the proposed BiLSTM structure, which has five layers: input, convolutional, full, LSTM, and output. The convolutional layer and pooling layer in the above layers have been applied to the CNN model and has proved effective for image recognition. Through the kernel, the convolution procedure reduces the input data to a smaller function map. Because the CNN input is a two-dimensional number, convolution kernels and maps are typically two-dimensional. To meet the one-dimensional characteristics of mechanical signals, we build in this paper a single-dimensional neural network of convolution, and all feature kernels and maps are one-dimensional.

Data is stored in the time domain and can be sorted chronologically, as indicated in the previous section. In order for the model to be particularly precise in time series processing, it must account both regional and global capabilities. A one-dimensional structure exists in time series. High correlation exists between variables (or pixels) that are near in time. The advantage of extracting and integrating regional features before recognising them is due to regional correlation. For local feature extraction, convolutional networks restrict the receptive field of hidden units. To handle global characteristics, LSTMs, on the other hand, can learn long-term dependencies between two entities. To handle global characteristics, LSTMs, on the other hand, can learn long-term dependencies between two entities. As a result, combining these two networks allows for reliable data processing. In general, compared to other temporal prediction algorithms, CNNs offer denoising qualities that lessen the impact of noise during training (LSTMs are noise sensitive) and require less preprocessing. Furthermore, compared to previous DNN models, the CNN-BiLSTM model successfully minimises overfitting.

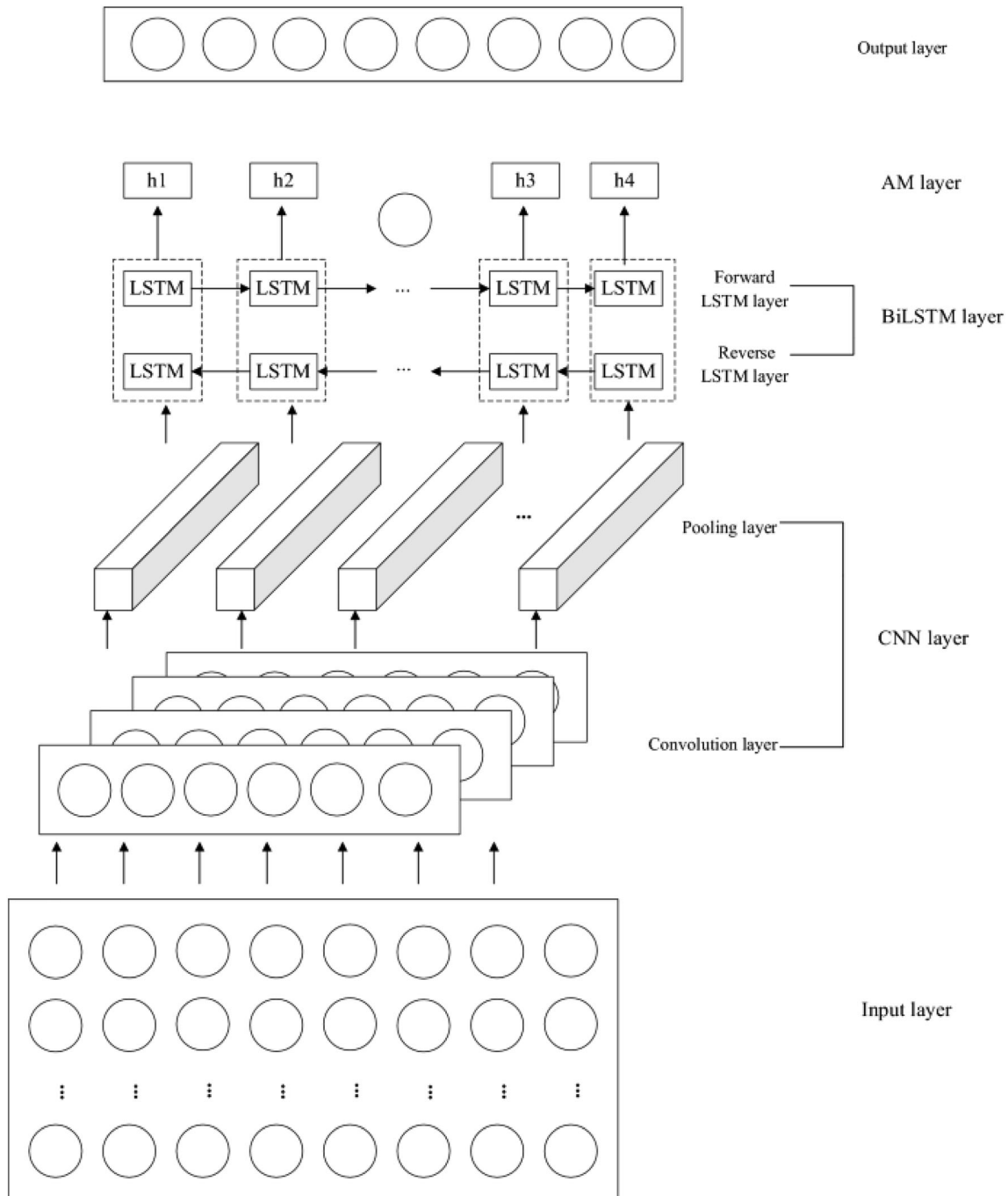


Figure 4. CNN with Bi LSTM Architecture.

The upgraded CNN-BiLSTM model is chosen for bearing problem diagnosis based on the following benefits and thorough experimentation to identify the optimum model to obtain the best accuracy in the shortest time period.

The CNN-BiLSTM architecture is depicted in figure 4. By minimising the cost function, model hyperparameters can be found (average loss function over the entire training set). CNN and BiLSTM layers make up the

suggested architecture. The proposed network evaluates the CWRU bearing dataset in order to identify the appropriate hyperparameters for minimising the cost function. Using an 84×84 -dimensional CNN and a 24-neuron LSTM, we were able to attain satisfactory accuracy across the tests. A dropout layer follows each main layer, thereby minimising overfitting by reducing the correlation between neurons.

Algorithm:

- Step 1:** Data preprocessing of corpus T_{train} , T_{test} ;
 - Step 2:** Get the set W of the fault labeled data;
 - Step3:** Initialize CNN–BiLSTM model parameters;
 - Step 4:** For each sample $t \in T_{train}$
 - Step 5:** The set of fault label as $t = w_1, w_2, \dots, w_n$
 - Step 6:** Obtaining eigen value $c_j = f(w) \times m + b$ by convolution operation
 - Step 7:** Compute eigenvector as $h = [h_1, h_2, \dots, h_n]$ through BiLSTM memory storage unit
 - Step 8:** Compute eigenvectors by $c_j^k = [(c_1, h_1), (c_2, h_2), \dots, (c_n, h_n)]$
 - Step 9:** maximum pooling value find by $c_j^k = \max(c_{i,j})$
 - Step 10:** To modify model parameters, a back propagation approach is used.
 - Step 11:** Using the Softmax function, calculate the likelihood of the fault ring label.
 - Step 12:** End for
 - Step 13:** For each fault data sample $t \in T_{train}$
 - Step 14:** Classification using trained CNN–BiLSTM model
 - Step 15:** End for
- Flow Diagram:**

Default data sets based on defect classification can generally be trained in smart fault diagnostic models. However, the model is inefficient or rather ineffective due to multiple part interactions or the effects of strong background noise when used for the initial weak diagnosis of composite parts and gearboxes. In addition, in real industrial applications, training data is generally limited on the state of other machines, especially in defective states. For example, it works under normal circumstances in the case of a gearbox. Therefore, training data is relatively difficult to accurately diagnose failures with millions of parameters, particularly as defect data is limited. The proposed solution is the same as in figure 5, and defects can be found by following these steps:

- (1) Vibration signals from the rolling bearing are collected by the capture device. Resampling is used to split the raw

- vibration data into training and testing samples. Transform wavelets are used to transform the raw vibration signals recorded by the gearbox into TF pictures.
- (2) Suggested extraction method for multi-scale features based on multi-scale units Get additional deep features through multiple layers of convolution and pooling layers.
- (3) The acquired deep characteristics are fed into the global mean complex hierarchy to achieve channel tuning. Then define the output with the softmax classifier, and train the source domain model with the BP algorithm.
- (4) Move the pre-trained model to different parts of the source area. Use a small sample of target fields with consistent weights in several units of measure to fine-tune the parameters of the delivery model. The CNN–BiLSTM model is trained using both the generated and original training picture samples.
- (5) The test image set is entered in a proposed model trained to detect rotary machine failure. Output is the result of failure detection.

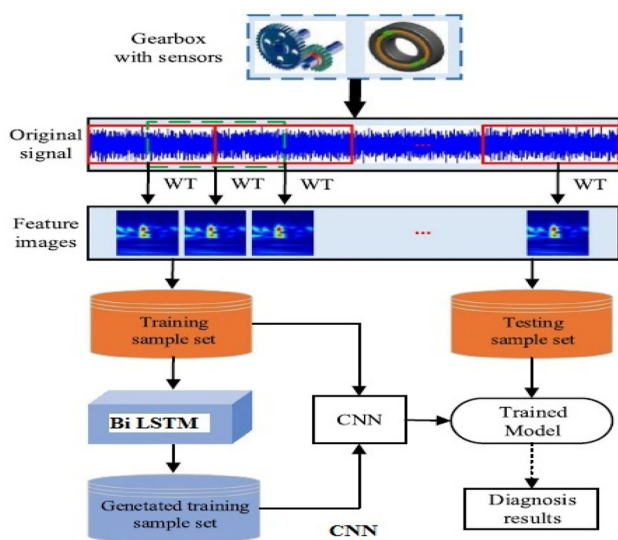


Figure 5. The flow diagram of the proposed method.

3.3 Training process

Figure 6 depicts the training process, in which the output classes are compared to the actual sample labels, and the parameters are changed using the BP algorithm. After numerous repetitions, the value of the loss function drops, suggesting that the parameter matches the sample characteristics. For determining the batch size of training data for iteration, the mini-batch slope-descent method is employed to speed up training and avoid region-optimal situations.

3.4 Testing process

The test dataset is added to the trained model after training is completed, as shown in figure 7. Use the modified dataset

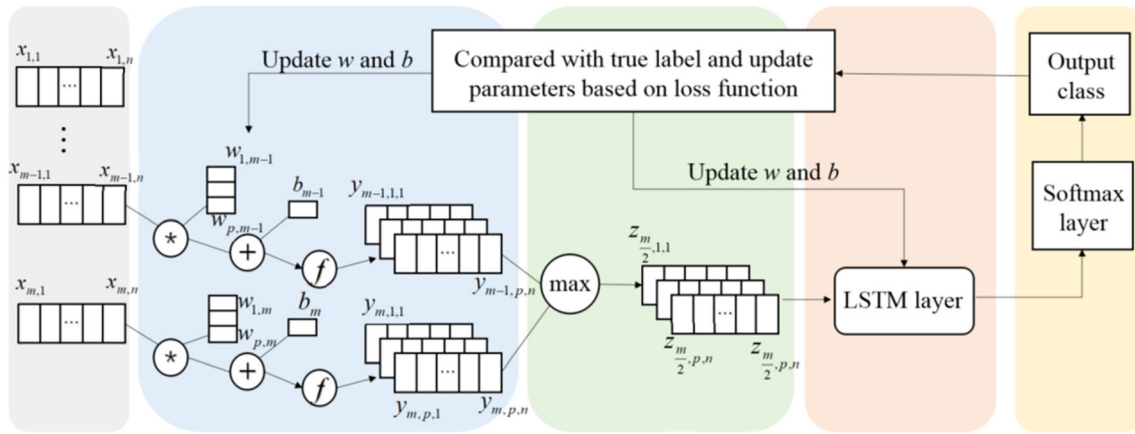


Figure 6. Training Process.

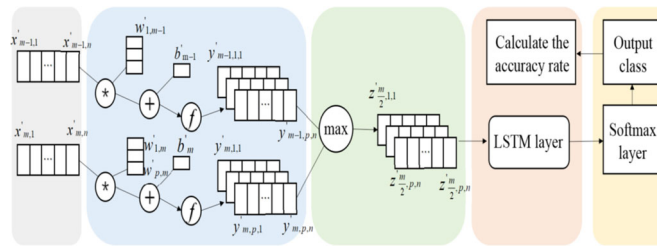


Figure 7. Testing Process.

settings from the previous training iteration. This is the flow chart for the whole model presented in Figure 7, where N represents the maximum iteration epoch. The vibration signal from the test bench is first collected and separated into motion and test datasets. Then, using the training dataset, we train the model several times to find the best parameters. Third, we feed the test data set into the trained model to obtain the predicted class, allowing us to calculate the accuracy using the formula.

$$Accuracy = \frac{\text{No. of predicted samples}}{\text{No. of total samples}} \quad (11)$$

4. Experimental results

The experimental dataset is based on the open source data of the data centre bearing the Case West Reserve University (CWRU). A standard database should be used to evaluate the performance of the proposed approach to that of existing methods. The CNN-BiLSTM is implemented using a Neural Network Toolbox in MATLAB 2018a, a CPU with Ryzen 5 1600, 16GB memory and GPU on the GTX 1060 computer.

4.1 Data description

Signals from driving stage bearings and EDM-created bearing problems are collected in this test (Electrical Discharge Machining). A mechanical motor drive system with sampling frequencies of 12 kHz and 48 kHz was used to obtain raw vibration data. Under varied loads, the testbench replicates four types of problems. Under various loads, the testbench mimics four fault categories. Stable, ball, inner ring, and outer ring faults are the four different varieties.

Figure 8 shows the platform for the collection, mainly made up of engines, torque sensors, power metres and electronic controls. With loads of 0, 1, 2, and 3 HP, the engines run at 1797, 1772, 1750, and 1730 rpm. 7 mil, 14 mil, 21 mil, and 28 mil are among the wear flaws. The vibration signal has sampling frequencies of 12K and 48 K when manually set.

4.2 12K CWRU dataset

In this study, we have used a 12 K sampling frequency, a 1797 rpm bearing speed and 0 HP drive termination signal. The conditions of bearings included normal orbital internal defects, external orbital defects and ball defects, with

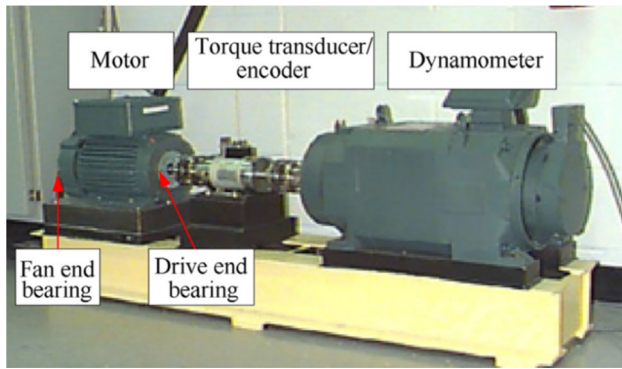


Figure 8. CWRU bearing fault test-bed.

normal (14-mils) and severe (11-mils) wear being small (7-mils). Therefore, there are 10 types of bearing conditions: normal condition (Normal), slight defect on the internal raceway surface (IR007), intermediate defect on the internal raceway surface (IR014), and serious defect on the internal raceway surface (IR021). Slight defect in external trajectory (OR007), intermediate defect in external trajectory (OR014), serious defect in external trajectory (OR021), slight defect in ball (B007), serious defect sphere with intermediate defect in ball (B014) At (B021). The size of the converted image is 160×160 , 374, and is transported from a sample signal to a screen at 25,600 sample points.

4.3 48K CWRU dataset

The rolling element fault signal, inner and outer rings, and normal signal are all displayed from top to bottom using the 48 KHZ drive termination signal. The driver must be carried during this test. The bearing to be diagnosed has a model number of SKF6205, and the sample frequency is 48 kHz. The balls, inner ring, and outer ring of the bearing are all susceptible to failure. 0.007, 0.014, and 0.021 inches, respectively, are the defect diameters. There are a total of ten different damage instances.

4.4 Data preprocessing

The drive bearing faults at the 48 K and 12 K ends of the CWRU data were preprocessed as shown in figures 9 and 10. The BiLSTM model's input layer is made up of $400 \times 20 \times 20$ -shaped neurons. For 2D convolution, the input data is first passed through the CNN network. The sliding stride is 2×2 , and the convolution kernel is 2×2 . External padding can contaminate the distribution of the data, but neither convolutional layers nor pooling layers contain embeddings to prevent this. Assume that the input data has a volume of H1L1D1 and that the convolutional layer has a size of H2L2D2. The CNN findings are then reproduced and processed by hierarchical attention layers in

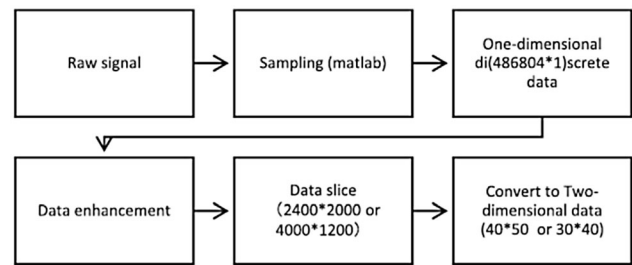


Figure 9. CWRU data preprocessing process.

order to acquire weighted resources while keeping the $8 \times 8 \times 32$ data dimension intact. The LSTM network's output can be combined with an entire connection layer to produce ten separate fault diagnosis results that softmax must alter.

For the CWRU 48 K and 12 K datasets, the accuracy of the proposed CNN-BiLSTM is compared to traditional approaches, with the best results listed in table 1. The results are shown that the efficacy of the proposed method, which by its density prediction function overcomes the comparative DL method. For all DL models in this study, CNN-BiLSTM achieved the highest precision of 99.80% for 48 K dataset. For 48 K dataset, CLSTM achieved accuracy of 98.33% and LSTM showed its accuracy performance of 95.44%. For CNN-BiLSTM achieved the highest accuracy of 98.30% for 12 K dataset. Similarly for 12 K dataset, CLSTM achieved accuracy of 95.7% and LSTM accuracy performance of 93.3%. Conversely, the CNN-BiLSTM model uses hierarchical network architecture to store useful vibration images information to improve performance than comparative models. CNN-BiLSTM achieved the highest sensitivity of 96.78% for 48 K dataset and 95.53% for 12 K dataset and also achieved the best specificity of 98.18% for 48 K dataset and 98.30% for 12 K dataset.

The accuracy of classification is 99.80%, much higher than the other 6 methods. The results of the experiments can draw the following conclusions: (1) Auxiliary data sets help improve the accuracy of classification and provide an effective way to diagnose missing labels error. (2) The proposed method has better stability for recognition of defects.

The simulation result of 48 K classification accuracy depicted in figure 11. Figure 12 depicts the test's CWRU 48 K dataset confusion matrix. Except for classes 2 and 8, the classifier accurately predicted almost all of the ten classifications.

The simulation result of 48 K classification accuracy depicts in figure 13. The CWRU 12 K dataset confusion matrix for the test is displayed in figure 14. Almost all 10 classes were correctly predicted by the classifier except class number 9.

The training accuracy and loss function of the DL model. This model has been trained to stop the training in

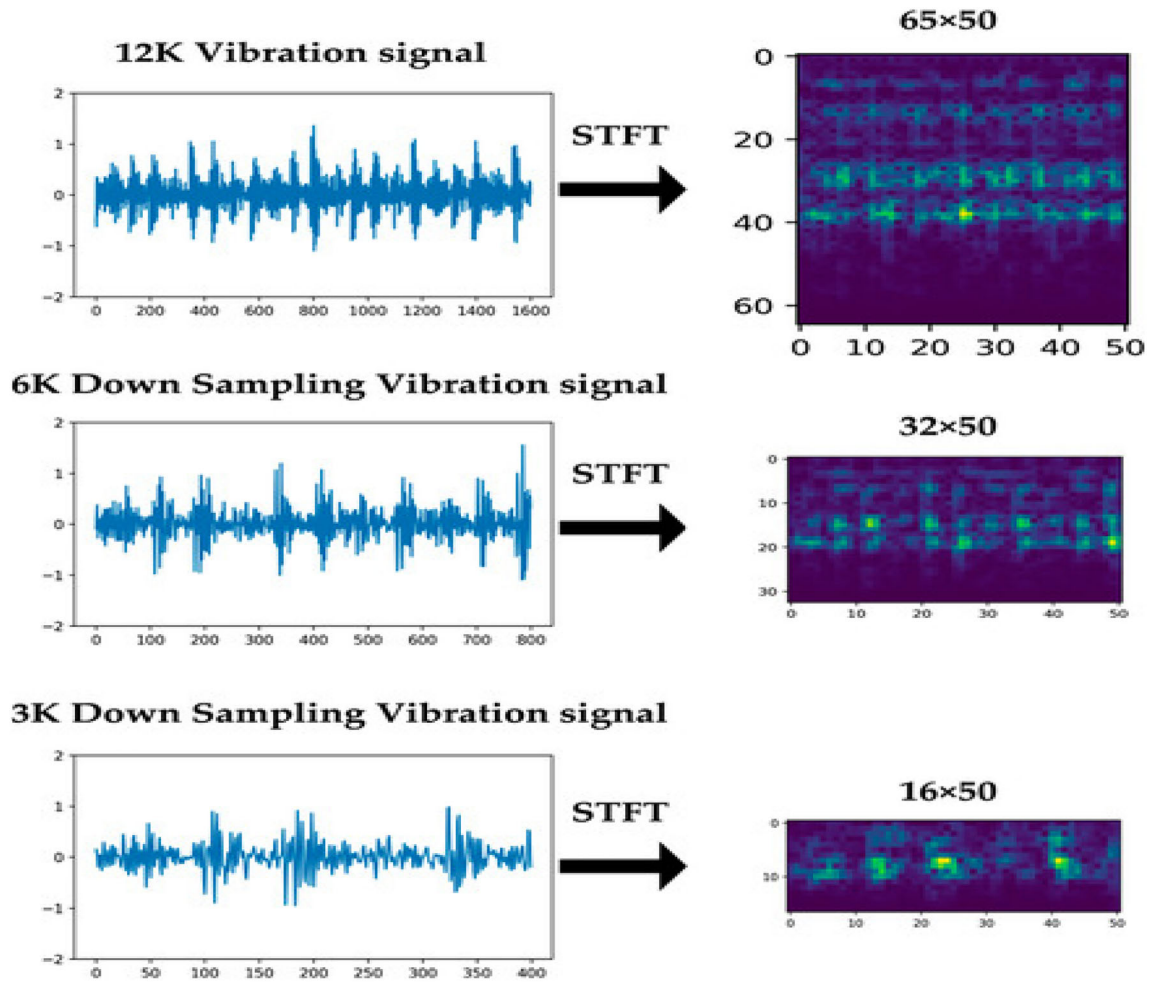


Figure 10. Data preprocessing of vibration signal.

Table 1. Fault diagnosis accuracy of different algorithms on CWRU 48 K and 12 K data set.

Methods	Accuracy (%)		Sensitivity (%)		Specificity (%)	
	48K	12K	48K	12K	48K	12K
FCN	77.78%	75.6%	74.56%	74.14%	76.68%	78.20%
SVM	59.44%	78.4%	60.43%	76.83%	64.32%	81.49%
Decision Tree	68.89%	77.6%	67.90%	77.95%	70.58%	84.19%
KNN	92.67%	85.8%	90.13%	83.58%	95.93%	87.72%
LSTM	95.44%	93.3%	92.45%	91.17%	98.32%	94.89%
CLSTM	98.33%	95.7%	93.29%	95.70%	97.48%	96.54%
Proposed	99.80%	98.3%	96.78%	95.53%	98.18%	98.30%

an early stop callback when the DL model achieves the highest precision without overfitting. The number of epochs therefore depends on each DL model. The BiLSTM model’s accuracy tends to remain stable following the 30th epoch, achieving an accuracy of 99.8%, and improving the loss function value after the 25th

epoch. The CLSTM model achieves 98.33% accuracy, but the model response continues to vary. Similarly, the LSTM, provides 95.44% accuracy in response to fluctuations in both accuracy and loss. The conventional KNN accuracy is approximately 92.71%, improving accuracy and loss.

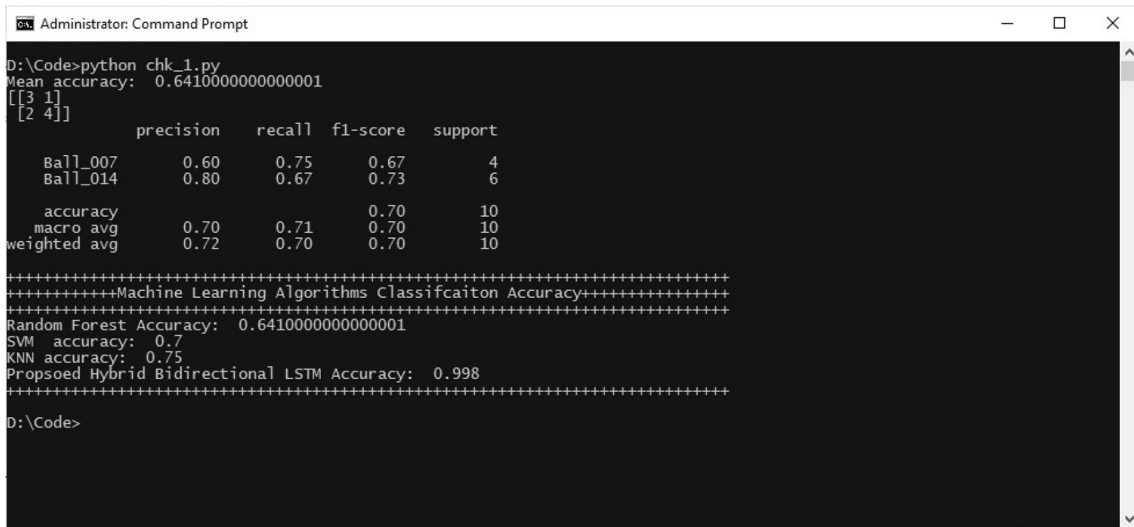


Figure 11. Simulation Result of Proposed Accuracy of CWRU 48 K Dataset.

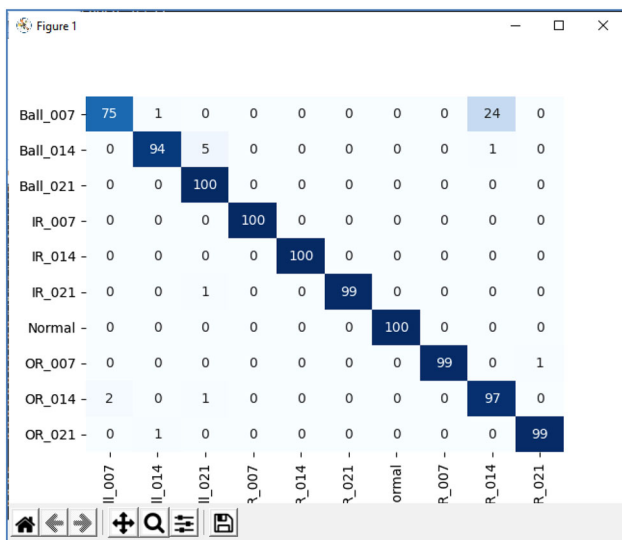


Figure 12. Confusion matrix of CWRU 48K dataset.

Randomly select 100 test samples and draw the predicted life curves of baseline, 48 K and 12 K dataset are as shown in figures 15, 16 and 17. The result is basically the same as the final result obtained repeatedly in the previous training. The results of the distribution graph are basically kept within a small error range. Therefore, the proposed CNN-BiLSTMbased bearing remaining life prediction method is effective, and the bearing life can be highly estimated. The predicted values are nearly identical at each selected location. Data points representing good RUL performance neural network.

Figures 18 and 19 show the results of RUL predictions using 48 K and 12 K dataset. Related errors are also shown in the chart. The technique satisfies the researched bearing

deterioration trend with minor error when comparing the anticipated RUL trend to the actual situation. To display RUL estimated from real bearing life, the indexed CNN-BiLSTM supervised model is utilized as a prediction tool. Experiment one uses CNN-BiLSTM depth features extracted from BiLSTM to perform RUL prediction. There is less histogram error in figure 18 when using the depth feature than when using the traditional statistical feature, which is shown to be closer to real-time. These comparisons show that CNN-BiLSTM is more obvious to the degradation process and can extract complete features. The initial trend in the forecast indicates that the actual value was initially approximated. Since BiLSTM uses randomly initialized neurons, as the sample size decreases, the randomness of prediction is inevitable. Repeat each model selection method 160 times to eliminate the negative effects of randomness. Bearings for offline training and testing were selected from a set of six. However, at the end of the bearing life, the predicted value will deviate from the actual value.

5. The effectiveness of CNN-BiLSTM classifier

The present CNN approach and the rolling bearing problem diagnosis method are compared and tested in the same settings in order to verify the effect of the method suggested in this study. Each approach was 300 times trained during the trial. Each time, 20 sites from 1260 training samples are chosen at random. That is, your network is trained using 640 samples. Following each training session, 32 samples were picked at random from the 420 test samples for network testing, and the recognition rate was calculated. The CNN-BiLSTM method has a greater and more consistent recognition rate than the conventional CNN method. This is due to the fact that standard CNNs are

```

Administrator: Command Prompt - python Run_12k.py
Epoch 35/40 [=====] - 3s 99ms/step - loss: 0.0295 - accuracy: 0.9906
Epoch 36/40 [=====] - 3s 89ms/step - loss: 0.0288 - accuracy: 0.9924
Epoch 37/40 [=====] - 3s 87ms/step - loss: 0.0325 - accuracy: 0.9889
Epoch 38/40 [=====] - 3s 86ms/step - loss: 0.0217 - accuracy: 0.9935
Epoch 39/40 [=====] - 3s 87ms/step - loss: 0.0413 - accuracy: 0.9835
Epoch 40/40 [=====] - 3s 87ms/step - loss: 0.0197 - accuracy: 0.9959
Iteration 10, Accuracy: 0.9570
Iteration 10, Accuracy: 0.9710
Iteration 10, Accuracy: 0.9790
Iteration 10, Accuracy: 0.9760
Iteration 10, Accuracy: 0.9620
Iteration 10, Accuracy: 0.9830
Iteration 10, Accuracy: 0.9790
Iteration 10, Accuracy: 0.9600
Iteration 10, Accuracy: 0.9710
Iteration 10, Accuracy: 0.9720
Iteration 10, Accuracy: 0.9570
[0.97100002 0.97899997 0.97600001 0.96200001 0.98299998 0.97899997
 0.95999998 0.97100002 0.972 0.95700002]
Mean Accuracy:0.9710
Best Accuracy: 0.9830
Worst Accuracy: 0.9570
Standard Deviation: 0.0083
    
```

Figure 13. Simulation Result of Proposed Accuracy of CWRU 12K Dataset.

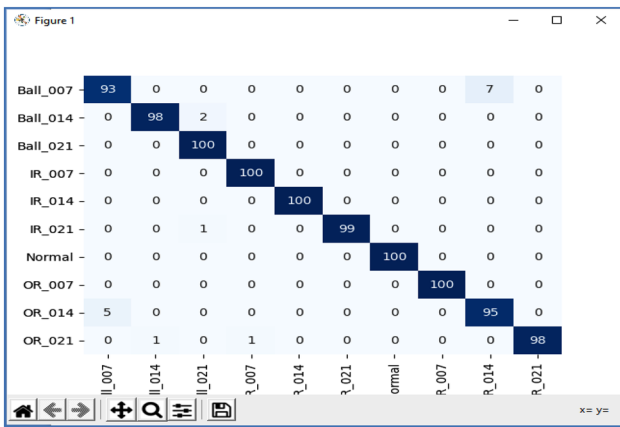


Figure 14. Confusion matrix of CWRU 12K dataset.

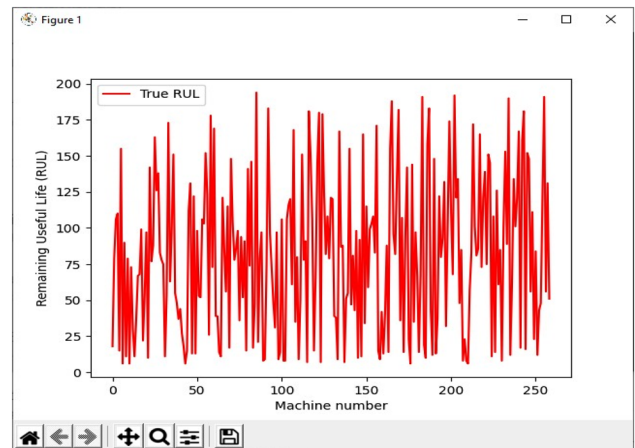


Figure 16. RUL results of 48 K dataset.

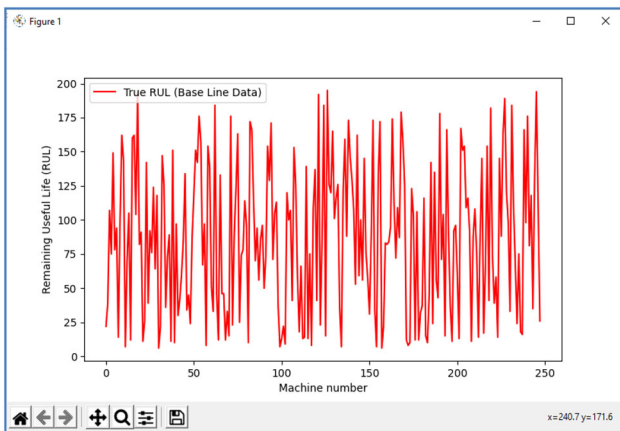


Figure 15. Remaining useful life (RUL) results of baseline dataset.

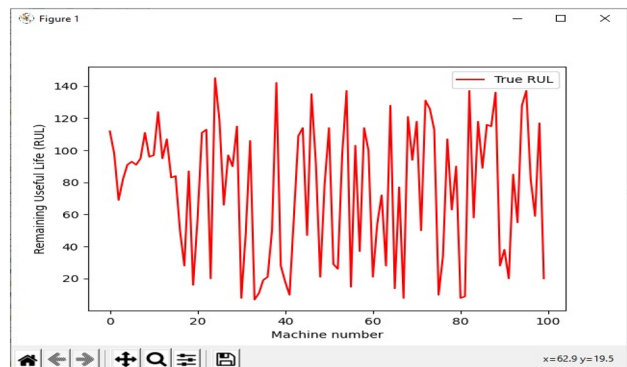


Figure 17. RUL results of 48 K dataset.

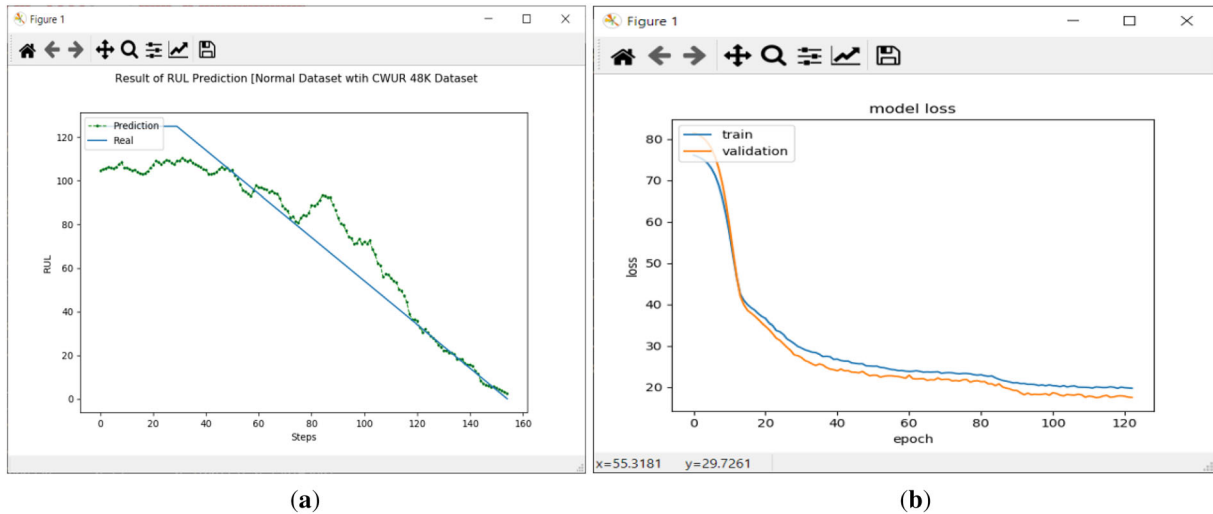


Figure 18. (a) RUL results of 48 K dataset and (b) Loss Model.

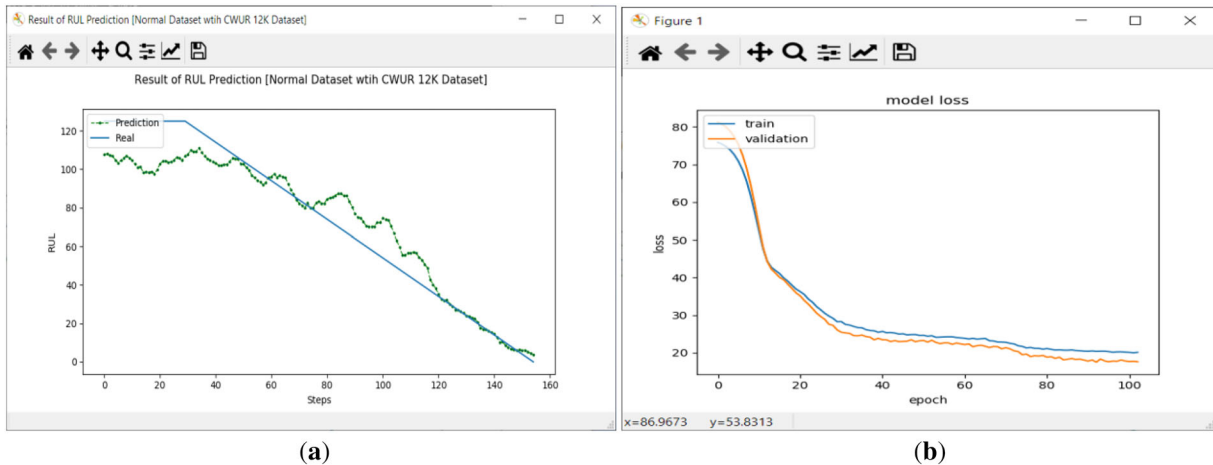


Figure 19. (a) RUL results of 12K dataset and (b) Loss Model.

constrained by fixed kernel shapes and lack adaptive mechanisms for converting geometry into complex pictures in real-time training.

By comparing them with LSTMs, CNN-BiLSTM generates random forests, SVMs, KNNs, FCNs, decision trees, and a variety of auxiliary datasets. To demonstrate the efficacy of CNN-BiLSTM as a classifier, the following conclusions may be taken from the experimental findings. (1) The auxiliary dataset enhances classification accuracy and makes diagnosing missing label errors more efficient. (2) For flaw detection, the proposed approach is very stable.

5.1 State-of-the-art techniques

For the CWUR 48 K dataset, table 2 demonstrates enhanced SVM schemes, commonly used UNET and DBN schemes, and signal processing and machine learning combinations like

CNN-SDAE, CNN-LSTM, and multi-scale CNN-LSTM. This paper’s proposed CNN-BiLSTM technique outperforms current methods in terms of diagnostic performance and offers a lot of room for improvement.

For the CWUR 12 K dataset, table 3 offers improved SVM approaches, commonly used DBN and SAE methods, and signal processing and machine learning combinations such as EEMD+AR+SVM, EMD+BP+LSFLA. The proposed CNN + BiLSTM method achieved highest accuracy as 98.3% among all the conventional methods.

6. Conclusion

Bearing failure can be diagnosed using CNN-BiLSTM in this article. Harmonics in the signal are first removed using a Fourier basis, while the impact and noise are preserved.

Table 2. State of art Techniques of CWRU 48 K dataset.

Authors	Methods	Accuracy
Dileep <i>et al.</i> , 2021	UNET	98.91%
Li and Zhang 2019	LSTM	96%
Honghu and Kann 2018	CNN	99%
Che and Peng 2020	CNN-SDAE	96.5%
Wen <i>et al.</i> , 2015	BP	86.4%
Eren <i>et al.</i> , 2017	DBN	89.4%
Wen <i>et al.</i> , 2019	SVM	83.7%
Zhu <i>et al.</i> , 2020	Multiscale CNN	98.46%
Proposed	BiLSTM	99.80%

Table 3. State of art Techniques of CWRU 12 K dataset.

Authors	Methods	Accuracy
Wang and Du 2012	Multi Class SVM	91.99%
Shao <i>et al.</i> , 2015	DBN	87.45%
Yuwono <i>et al.</i> , 2016	PSO with HMM	88.02%
Hoang DT and Kang, 2018	EMD + BP	89.33%
Chen and Peng, 2020	H-CNN	92.60%
Qi <i>et al.</i> , 2017	EEMD+AR+SVM	94.65%
Li <i>et al.</i> , 2020	GDCCN	94.62%
Goyal <i>et al.</i> , 2020	DWT + SVM	95.80%
Proposed	CNN + BiLSTM	98.3%

DWT is used to analyze the signal with shock and noise components to obtain a time-frequency map. Finally, the 10-layer CNN has a deformable convolutional kernel to diagnose bearing errors. Experimental results show that the vibration signal time frequency diagram after harmonic removal shows that the vibration signal characteristics are clear and can be used as a CNN input to achieve satisfactory classification results. By adding a layer of hierarchy to each neural network sample, this method improves the CNN-BiLSTM method. These improvements improve the model's performance and achieve 99.95% of 48 K and 98.3% of 12 K accuracy. We have shown that CNN may successfully increase roller-bearing failure recognition rate and stability by comparing Random Forest, SVM, KNN, FCN, Decision Tree, and LSTM on the same data set.

In the real world and industry, the implementation of the proposed method has the following advantages over other methods:

- Models are far too strong to keep track of massive volumes of information. In comparison to other literatures, you can create more accurate forecasts in a shorter amount of time.
- Data preprocessing does not require any transformation, modification, feature extraction, or feature

selection because real-time vibration data can be delivered immediately.

- In comparison with some solutions, computations are cost effective, including preprocessing data and the complex, profound architectures described in the literature.
- The best accuracy achieved as 98.3% for 12 K and 99.8% for CWRU 48 K bearing data sets.
- According to the results, predicting RUL bearings using the proposed method is more accurate than using a traditional method.

6.1 Future work

The proposed CNN-BiLSTM neural network topology in this paper does not solve the RNN parallel computing problem fundamentally. Furthermore, several flaws are not processed simultaneously. These are the positions that we shall have in the future. Furthermore, several flaws are not processed simultaneously. Because cloud-based data has a simple logical structure and no grammatical logical relationship between data, this is the case. As a result, identifying defects or extracting crucial data from nearby data points is impossible.

Abbreviations

DL	Deep learning
CNN	Convolutional neural networks
CWRU	Case Western Reserve University
ANN	Artificial neural networks
PCA	Principal component analysis
AI	Artificial intelligence
FMLE	Fusel maximum likelihood optimization
LSTM	Long short term memory
BiLSTM	Bidirectional LSTM
WHN	Weighted hierarchical networks

Declaration

Conflict of interest The authors declare that they have no conflict of interest. Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

- [1] Chen X, Zhang B and Gao D 2020 Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J. Intell. Manuf.* <https://doi.org/10.1007/s10845-020-01600-2>
- [2] Wu K, Wu J, Feng L, Yang B, Liang R, Yang S and Zhao R 2020 An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system. *Int. Trans. Electr. Energ. Syst.* 31. <https://doi.org/10.1002/2050-7038.12637>

- [3] Zhou F, Yang S, Fujita H, Chen D and Wen C 2020 Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl. Based Syst.* 187: 104837. <https://doi.org/10.1016/j.knosys.2019.07.008>
- [4] Pan H, He X, Tang S and Meng F 2018 An improved bearing fault diagnosis method using one-dimensional CNN and LSTM. <https://doi.org/10.5545/sv-jme.2017.5249>
- [5] Eren L, Ince T and Kiranyaz S 2018 A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *J. Signal Process. Syst.* <https://doi.org/10.1007/s11265-018-1378-3>
- [6] Zou P, Hou B, Jiang L and Zhang Z 2020 Bearing fault diagnosis method based on EEMD and LSTM. *Int. J. Comput. Commun. Control* 15(1): 1010. <https://doi.org/10.15837/ijccc.2020.1.3780>
- [7] Yink W and Kann K 2017 Comparative study of CNN and RNN for natural language processing. *Comput. Sci.* 1702(01923): 2017
- [8] Wu C and Jiang P 2019 Intelligent fault diagnosis of rotating machinery based on one dimensional convolutional neural network. *Comput. Ind.* 108: 53–61
- [9] Wang J and Du G 2020 Fault diagnosis of rotating machines based on the EMD manifold. *Mech. Syst. Signal Process* 135: 106443
- [10] Goyal D, Choudhary A, Pabla B S and Dhami S S 2020 Support vector machines based non-contact fault diagnosis system for bearings. *J. Intell. Manuf.* <https://doi.org/10.1007/s10845-019-01511-x>
- [11] Hoang D T and Kang H J 2018 Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cogn. Syst. Res.* 53: 42–50. <https://doi.org/10.1016/j.cogsys.2018.03.002>
- [12] Hochreiter S and Schmidhuber J 1997 Long short-term memory. *Neural Comput.* 9(8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [13] Jinde Z, Zhanwei J, Ziwei P and Kang Z 2016 VMD based adaptive multiscale fuzzy entropy and its application to rolling bearing fault diagnosis. In: *International Conference on Sensing Technology. IEEE.* <https://doi.org/10.1109/ICSensT.2016.7796267>
- [14] Lau E C C and Ngan H W 2010 Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis. *IEEE Trans. Instrum. Meas.* 59(10): 2683–2690. <https://doi.org/10.1109/TIM.2010.2045927>
- [15] Li J, Li X, He D and Qu Y 2020 Unsupervised rotating machinery fault diagnosis method based on integrated SAE–DBN and a binary processor. *J. Intell. Manuf.* <https://doi.org/10.1007/s10845-020-01543-8>
- [16] Liang Y, Li B and Jiao B 2020 A deep learning method for motor fault diagnosis based on a capsule network with gate-structure dilated convolutions. *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-020-04999-0>
- [17] Lu C, Wang Z and Zhou B 2017 *Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification.* Elsevier Science Publishers B. V. <https://doi.org/10.1016/j.aei.2017.02.005>
- [18] Qi Y, Shen C, Wang D, Shi J and Zhu Z 2017 Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *IEEE Access.* <https://doi.org/10.1109/ACCESS.2017.2728010.17>
- [19] Shao H, Jiang H, Zhang X and Niu M 2015 Rolling bearing fault diagnosis using an optimization deep belief network. *Meas Sci Technol.* 26. <https://doi.org/10.1088/0957-0233/26/11/115002>.
- [20] Shao H D, Jiang H K and Zhao H W 2017 A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* 95: 187–204. <https://doi.org/10.1016/j.ymsp.2017.03.034>
- [21] Song L, Wang H and Chen P 2018 Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery. *IEEE Trans. Instrum. Meas* 67(1887–1899): 2018
- [22] Wang Y, Kang S, Jiang Y, Yang G, Song L and Mikulovich V I 2012 Classification of fault location and the degree of performance degradation of a rolling bearing based on an improved hyper-sphere-structured multi-class support vector machine. *Mech. Systems Signal Process.* 29: 404–414. <https://doi.org/10.1016/j.ymsp.2011.11.015>
- [23] Wen L, Gao L and Li X 2017 A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Trans. Syst. Man Cybern. Syst.* <https://doi.org/10.1109/TSMC.2017.2754287>
- [24] Wen L, Li X, Gao L and Zhang Y 2017 A new convolutional neural network based data-driven fault diagnosis method. *IEEE Trans. Ind. Electron.* <https://doi.org/10.1109/TIE.2017.2774777>
- [25] Eren L, Ince T and Kiranyaz S 2019 A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *J. Signal Process. Syst.* 91(2): 179–189
- [26] Zhu Z, Peng G, Chen Y and Gao H 2019 A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis. *Neurocomputing* 323: 62–75
- [27] Prakash G, Narasimhan S and Pandey M D 2019 A probabilistic approach to remaining useful life prediction of rolling element bearings. *Struct. Health Monit.* 18(2): 466–485
- [28] Cui L, Huang J, Zhang F and Chu F 2019 HVSRMS localization formula and localization law: localization diagnosis of a ball bearing outer ring fault. *Mech. Syst. Signal Process.* 120: 608–629
- [29] Li H and Zhang Q 2018 Fault diagnosis method for rolling bearings based on short-time Fourier transform and convolution neural network. *J. Vib. Shock* 37(19): 125–131