Journal of Systems Engineering and Electronics
Vol. 34, No. 4, August 2023, pp.1074 – 1084

Rolling bearing fault diagnostics based on improved data augmentation and ConvNet

KULEVOME Delanyo Kwame Bensah [1,2], WANG Hong [1,2,*], and WANG Xuegang [1]

1. School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. Yangtze Delta Region Institute (Huzhou), University of Electronic Science and Technology of China, Huzhou 313001, China

Abstract: Convolutional neural networks (CNNs) are well suited to bearing fault classification due to their ability to learn discriminative spectro-temporal patterns. However, gathering sufficient cases of faulty conditions in real-world engineering scenarios to train an intelligent diagnosis system is challenging. This paper proposes a fault diagnosis method combining several augmentation schemes to alleviate the problem of limited fault data. We begin by identifying relevant parameters that influence the construction of a spectrogram. We leverage the uncertainty principle in processing time-frequency domain signals, making it impossible to simultaneously achieve good time and frequency resolutions. A key determinant of this phenomenon is the window function's choice and length used in implementing the short-time Fourier transform. The Gaussian, Kaiser, and rectangular windows are selected in the experimentation due to their diverse characteristics. The overlap parameter's size also influences the outcome and resolution of the spectrogram. A 50% overlap is used in the original data transformation, and ±25% is used in implementing an effective augmentation policy to which two-stage regular CNN can be applied to achieve improved performance. The best model reaches an accuracy of 99.98% and a cross-domain accuracy of 92.54%. When combined with data augmentation, the proposed model yields cutting-edge results.

Keywords: bearing failure, short-time Fourier transform, prognostics and health management, data augmentation, fault diagnosis.

DOI: 10.23919/JSEE.2023.000109

1. Introduction

Increased sensor integration in systems to provide a more comprehensive, reliable, and up-to-date acquisition of information on their operation status results from a need for improved reliability. Vibration analysis has been widely used for condition monitoring and defect diagnostics in rotating machinery over the past few years. An effective and efficient signal processing technique is essential in extracting characteristic features from the raw monitored signal to determine the severity of the damage to enable accurate modeling of a prognostics and health management (PHM) framework. Over the years, researchers have used various techniques to analyze the time, frequency, and time-frequency domain characteristics [1−3] of a vibration signal to determine the condition of bearings. Such traditional approaches require ample prior knowledge of fault detection and signal processing, which is inconvenient for industrial applications. As a result, intelligent fault diagnostics techniques have become increasingly common, which can produce accurate fault diagnostics outcomes with no previous experience. These data-driven techniques are formulated using machine learning (ML) models to classify fault types by extracting significant features.

In recent years, the advancement and implementation of deep learning (DL) in component diagnostics, prognostics [4,5] and image classifications [6−8] has motivated research in the application of such models in bearing fault diagnostics [9,10]. This research aims to leverage the capabilities of a convolutional neural network (CNN), which comprises sequentially composed layers of convolution and pooling operations, in establishing an effective fault classification model.

Researchers adopt different methods of preprocessing the vibration data used in classification models. A direct translation approach entails a conversion of segments of

Manuscript received May 12, 2022.
*Corresponding author.
This work was supported by the National Natural Science Foundation of China (42027805), and the National Aeronautical Fund (ASFC-20172080005).

the raw signal into matrices [11]. Similarly, wavelets [12] and spectrogram [10] are other representations of the raw vibration input signals. Some successes have been made in end-to-end diagnostic models that can explicitly bridge the connection between raw monitoring data and system health states. However, they are primarily based on the premise that labeled data are adequate and provide full knowledge about machine health states. However, in engineering scenarios, such an inference is infeasible since these data cannot provide enough detail to represent all of the different types of health conditions [13].

Machines operate in a stable environment, with defects occurring only infrequently. As a result, the accumulation of healthy data is more prominent than unhealthy data leading to an imbalance in a vast majority of the data gathered. Since it is challenging to accrue sufficient operating data needed to implement a DL model, data augmentation (DA) techniques can address this situation by artificially creating new samples based on existing training data. This reduces the effect of overfitting and improves the model's efficiency. However, despite the successful implementation of DA in other applications, not much research has been done in machine health diagnostics tasks. The addition of noise [14] and using generative models [15−17] to generate synthetic datasets are some existing methods, with the latter being prominent.

This paper proposes new DA techniques for fault diagnostics of bearings based on a simple but efficient CNN model. The main contributions of this research are summarized as follows:

(i) Propose DA that utilizes the various window functions to enhance either time or frequency resolution of spectrogram representations of bearing operating conditions, and adopts different lengths of overlaps between segments in spectrogram generation.

(ii) Propose the implementation of non-negative matrix factorization (NMF) to generate more training data samples to circumvent data insufficiency.

(iii) Perform several DA on the training set to enhance the generalizability of the classification model.

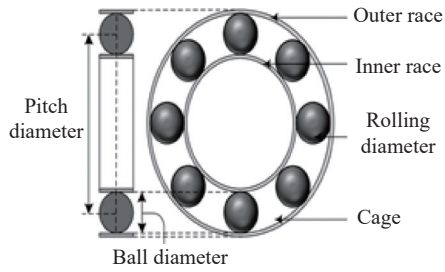
(iv) An efficient bearing fault diagnosis is presented by combining the proposed augmentation techniques and a lightweight CNN structure balancing low computation time and accurate fault classification.

(v) Perform extensive experimental evaluation using a publicly accessible dataset and compare the proposed techniques with different state-of-the-art methods.

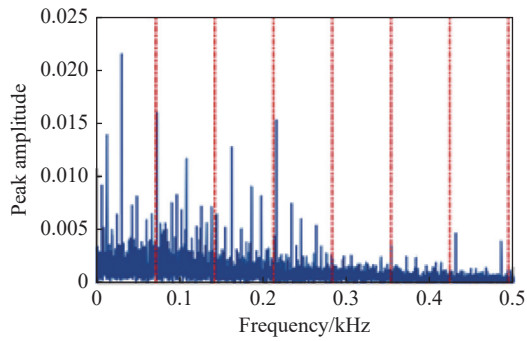
The remainder of this paper begins with a theoretical overview of the research in Section 2. In Section 3, a detailed description of our proposed methods is outlined, followed by experimental procedure and validation of the proposed methods in Section 4. The experiment results are presented and discussed in Section 5. The conclusion of this paper is drawn in Section 6.

2. Theoretical basis

Defects in bearings produce asynchronous vibration components in the inner and outer races, cage, and rolling elements. Depending on the geometry and speed of the bearing, each of these defects will produce a distinct frequency. These are the ball spin frequency (BSF), ball pass frequency inner (BPFI), and ball pass frequency outer (BPFO) [18]. These oscillations repeatedly occur with a period 'T'. If this coincides with one of the frequencies calculated from the bearing's geometry, then the location of a defect can be determined. A typical geometry and defects that can occur in a rolling bearing are shown in Fig. 1. The frequency indicates the source of the fault, while the amplitude signifies the severity of the fault. Vibration readings above the standard threshold indicate the emergence of a fault and deterioration of a bearing which can escalate and lead to severe equipment damage.



(a) Geometry



(b) Ball

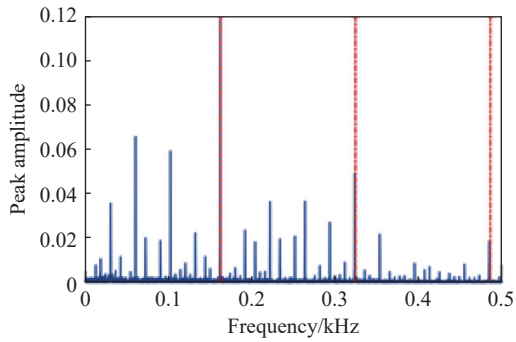


Fig. 1 Geometry and fault characteristics of a rolling element bearing

2.1 Short-time Fourier transform (STFT) for signal preprocessing

Vibration signals acquired from rotating machinery are non-stationary [18], making it difficult to describe and predict the relevant information within such signals based on their temporal or spectral representations.

A pragmatic approach is the time-frequency analysis by transforming the one-dimensional (1D) signal into a two-dimensional (2D) representation. A 2D data presents more vital information and subsumes complex structure distributions, enabling the prominence of intricate patterns that characterize a machine's operating condition.

The STFT is a linear transformation of a 1D signal to a 2D plane using a fixed-length sliding window. The STFT of a time-domain signal $x(t)$ can be expressed as

$$\mathrm{STFT}_x(t,f)=\int_{-\infty}^{+\infty}x(\tau)w(\tau-t)\mathrm{e}^{-\mathrm{j}2\pi f\tau}\mathrm{d}\tau \tag{1}$$

where $w(t-\tau)$ is the sliding window.

A spectrogram is the energy density spectrum of the STFT and can be obtained by taking the squared magnitude of the STFT expressed as

$$|\mathrm{STFT}_x(t,f)|^2=\left|\int_{-\infty}^{+\infty}x(\tau)w(\tau-t)\mathrm{e}^{-\mathrm{j}2\pi f\tau}\mathrm{d}\tau\right|^2. \tag{2}$$

Therefore, in this paper, the spectrograms are generated by carefully selecting optimum parameters resulting in fault classes with distinctive discriminative features.

2.2 DA

DL models require extensive labeled data during training to improve generalization capabilities. In case of insufficient training samples, DA is applied in the data preprocessing stage to make the model more robust. Also, it is often challenging to cover the entire sample space during data collection. Hence it is essential to incorporate an appropriate form of DA when training a model. Random geometric transformations are the most common types of DA that make assumptions about the underlying dataset's patterns [19]. However, time-series data have different properties from regular images, so these methods do not work for all of them.

2.3 CNN

CNN or ConvNet is a multi-stage neural network that comprises a feature learning and classification stage. The former consists of convolution, batch normalization, activation, and pooling layers. Multiple filters are convolved with the input data to generate feature maps in the convolutional layers. Subsequently, pooling layers extract the most significant local features. CNNs can learn abstract spatial features by alternating and stacking convolutional kernels and performing pooling operations. However, increasing the network's depth increases the computation time, which is undesirable. The classification stage is a multi-layer perceptron, which is composed of fully connected layers. For an input of size $M\times M$, having F number of filters of size $K\times K$ placed in L layers, the evaluation cost of the CNN architecture becomes $O(M^2K^2R^2L)$ [20]. The network storage is conditioned on the filters' size and biases, which results in $O(K^2R^2L)$. The computation complexity of training increases proportionally to the number of network variables and the size of the training dataset. When all the feature maps are included in the process, the memory complexity increases to $O(M^2RL)$ [20].

Devising faster ways of performing convolution without compromising the model's accuracy is highly desirable. In this research, we choose to get the most performance with few layers as possible by carefully selecting the model's hyperparameters.

3. Proposed method

The overall framework of the proposed techniques is shown in Fig. 2, which illustrates the stages involved in implementing the bearing operating condition diagnostics model.

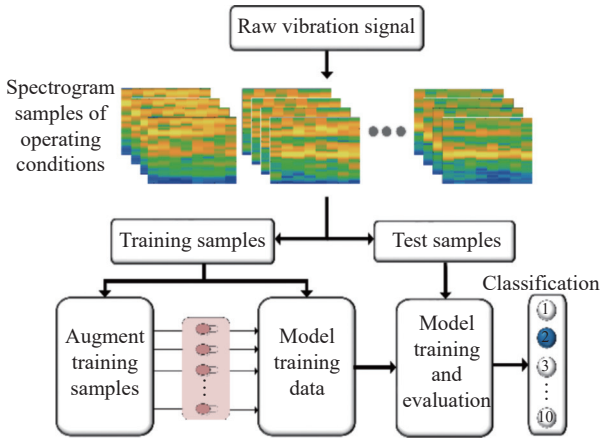


Fig. 2 Overall framework of the proposed approach

The raw vibration signals are first segmented and transformed into spectrograms. These are then divided into training and test sets. The generated training samples using the proposed augmentation methods are combined with the original training set to train the proposed CNN model. Finally, the performance of the trained model is evaluated by using test sample sets.

3.1 Proposed signal preprocessing

The raw vibration data samples are divided into equal samples with the same number of points by calculating the number of sample points N_s in one revolution. Therefore, given a sampling frequency F_s of 12 kHz, and an operating speed s of 1 797 rpm, the sample points per revolution can be estimated as

$$N_s = 60\frac{F_s}{s} = 60\frac{12\,000}{1\,797} \approx 400. \tag{3}$$

The total length of a vibration signal under all the faulty conditions is 120 000 for all operating loads. The length of the raw signal for a normal operating condition under 0 hp loads L_0 is 240 000, while that of 3 hp ($L_{1,2,3}$) is 480 000. Hence, demarcating the raw signal into samples of 400 data points as shown in Fig. 3 results in a total of 300 samples for all fault conditions under 0 hp, 1 hp, 2 hp, and 3 hp.

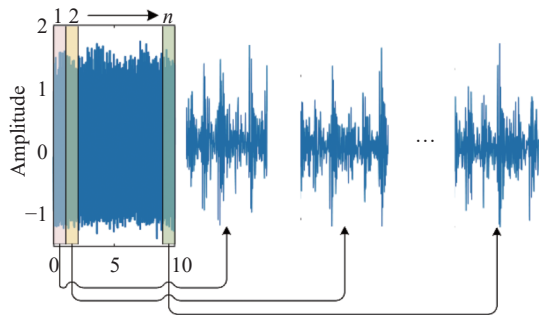


Fig. 3 An example illustrating the segmentation of an inner race fault signal into *n* samples

For normal condition, 600 samples are obtained for L_0 and 1 200 for $L_{1,2,3}$. Each of the samples is transformed into a spectrogram for efficient feature learning by the proposed CNN. The parameters used to compute a signal's spectrogram can significantly impact a classifier's performance [21]. Since there are virtually infinite combinations of such parameters, a section length of 88, overlap of 50%, 256 discrete Fourier transform (DFT) points, and a Hamming window are some of the optimum parameters selected in generating the spectrogram.

3.2 Proposed DA

3.2.1 Nonnegative matrix factorization

Given a nonnegative data matrix $\boldsymbol{X} \in \mathbf{R}_+^{n\times m}$, NMF aims to approximate $\boldsymbol{X}$ into a basis matrix $\boldsymbol{W} \in \mathbf{R}_+^{n\times r}$ and an encoding matrix $\boldsymbol{H} \in \mathbf{R}_+^{r\times m}$, where $\boldsymbol{W}$ and $\boldsymbol{H}$ are two low-rank nonnegative matrices. The factorization rank parameter r sets the number of factors to be used to describe the data and has a significant role in the process of factorization. Different values of r contribute to varying effects of factorization. $\boldsymbol{W}$ and $\boldsymbol{H}$ are obtained by applying an NMF to each sample of the bearing operating condition. The optimization problem is then formulated as

$$\min_{\boldsymbol{W}\in\mathbf{R}_+^{n\times r},\boldsymbol{H}\in\mathbf{R}_+^{r\times m}} f(\boldsymbol{W},\boldsymbol{H}) = \frac{1}{2}\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\|_{\mathrm{F}}^2 \tag{4}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix.

Since the optimization problem in (4) is nonconvex for both $\boldsymbol{W}$ and $\boldsymbol{H}$, the convex subproblem is realized by dividing the problem into two. A closed-form solution is obtained by using multiplicative update (MU) rule [22], alternating least squares (ALS) [23], or other gradient descent methods over a given number of iterations until convergence is reached. A local minimum may have functional characteristics in practice. Therefore, in this research, an excellent global minimum is not imperative. The steps used in implementing the NMF-DA is given in Algorithm 1.

Algorithm 1 Proposed pseudo-code of NMF-DA

Input: N is the number of classes in dataset; p is the percentage of overlap; w is the Hamming window; $\boldsymbol{W}^{(0)}, \boldsymbol{H}^{(0)} \in \mathbf{R}_+$ are random initial values; r is the rank of approximation; L is the length of a section; f_s is the sampling frequency

Output: $\boldsymbol{D}_{\mathrm{nmf}}$ is the NMF spectrogram

$\boldsymbol{D}_{\mathrm{nmf}} \leftarrow$ empty;

$m \leftarrow \max(256, 2^L)$;
for $c \leftarrow 1$ to N **do**
  sample $\boldsymbol{D}_c \in \boldsymbol{D}$ from dataset;
  foreach class data instance $\boldsymbol{x}_i \in \boldsymbol{D}_c$ **do**
    $[\boldsymbol{S}, \boldsymbol{f}, \boldsymbol{t}] \leftarrow$ computeSpectrogram$(\boldsymbol{x}_i, w, p, m, f_s)$;
    $\boldsymbol{X} \leftarrow$ Imagesc$(\boldsymbol{t}, \boldsymbol{f}, \boldsymbol{S})$;
    while a stopping criterion is not met **do**
      foreach $\boldsymbol{w}_r \; of \; \boldsymbol{W}$ **do**
        $\boldsymbol{w}_r \leftarrow \boldsymbol{w}_r / \|\boldsymbol{w}_r\|_2$;
      end
      $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{WH}$;
      for $r \leftarrow 1$ to R **do**
        $\boldsymbol{X}^{(r)} \leftarrow \boldsymbol{E} + \boldsymbol{w}_r \boldsymbol{h}_r$;
        $\boldsymbol{h}_r \leftarrow [\boldsymbol{X}^{(r)} \boldsymbol{w}_r]_+$;
        $\boldsymbol{w}_r \leftarrow [\boldsymbol{X}^{(r)} \boldsymbol{h}_r]_+$;
        $\boldsymbol{w}_r \leftarrow \boldsymbol{w}_r / \|\boldsymbol{w}_r\|_2$;
        $\boldsymbol{E} \leftarrow \boldsymbol{X}^{(r)} - \boldsymbol{w}_r \boldsymbol{h}_r$;
      end
    end
    $\tilde{\boldsymbol{X}} = \boldsymbol{WH}$
    $\boldsymbol{D}_{\text{nmf}_i} \leftarrow \log(\tilde{\boldsymbol{X}})$
  end
end
return $\boldsymbol{D}_{\text{nmf}}$

3.2.2 Adjusting relevant spectrogram parameters

(i) Localization trade-off

The resolution of the vibration signal's time-frequency domain (TFD) representation tends to be deficient in a particular domain depending on the chosen parameters. This is due to the uncertainty principle in processing the spectrogram where a perfect time and frequency localization cannot be achieved simultaneously [24]. This implies that a higher time resolution can be achieved for a given window at the expense of frequency resolution and vice versa.

We compose a DA strategy that generates variants of the input signal by utilizing the rectangular, Gaussian and Kaiser window functions [25]. For a given sample of raw vibration signal, the spectrogram is generated by adjusting parameters (α = 3, and β = 15) for Gaussian and Kaiser window functions. All other parameters remain constant, as detailed in Subsection 3.1.

(ii) Overlap

Each segment shown in Fig. 3 is further divided into sections of equal lengths in computing the spectrogram. The overlap parameter is essential in this process as it determines the interval at which the window hops over successive sections. This process results in the stretching of the time scale, which causes the changes in frequency with respect to time to be more visible. However, the overlap length between sections must be less than that of the window function. Therefore, 25% and 75% overlaps are selected to implement the proposed augmentation technique.

Examples of generated training samples using the proposed augmentation techniques are shown in Fig. 4.

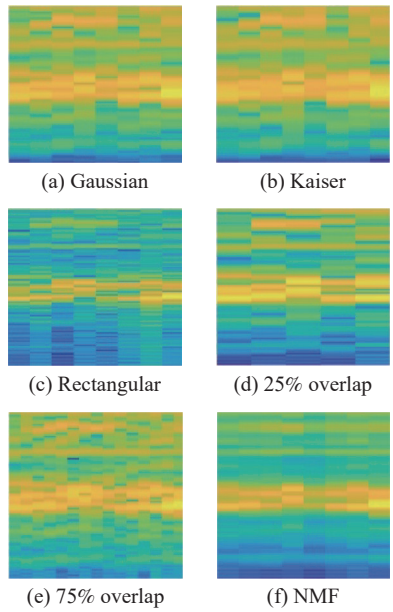


Fig. 4 Samples of the proposed DA techniques

3.2.3 Proposed CNN architecture

The 2D spectrogram data are reshaped to equal sizes of 32×32 and fed to the input layer of the model in mini-batches. Feature maps are generated through convolution operations using a 3×3 filter size at a stride of 1. The rectified activation function (ReLU) is implemented at this stage, followed by a batch normalization (BN) layer to reduce internal covariance shift, increase the learning process, and improve the network's overall performance. Subsequently, a pooling layer subsample the feature maps with a window of size 2×2 to reduce the spatial size. A max-pooling operation produces the maximum value in the 2×2 input region. The convolution, BN, and max-pooling operations are repeated to further extract intricate features. The outputs from the previous layer connect to a fully connected layer comprising two hidden layers with 64 units each and an output layer of size 10. The proposed architecture for diagnosing a bearing's operating condition is depicted in Fig. 5.

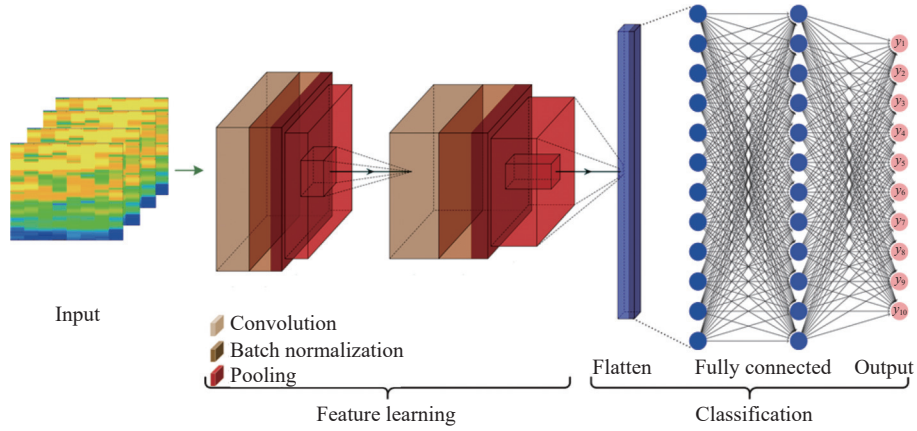


Fig. 5 Proposed CNN architecture for bearing fault diagnostics

The proposed architecture produces fewer trainable parameters and achieves high performance by adaptively capturing discriminative features from input data via multiple non-linear transformations. The parameters in the network are updated by using the back-propagation (BP) algorithm, and the Adam optimization approach is utilized to minimize the objective with batch size of 32. The suggested network's loss generally converges after ten epochs of training with a learning rate of 0.001. The trained model is finally used to test the remaining unseen samples to obtain the model's performance accuracy.

4. Experimental setup

The experimental setup of the proposed method on rolling bearing datasets is discussed in this section. The model's performance is thoroughly investigated with various DA techniques and combinations. The effect of randomness is minimized by averaging over five experimental trials, and the mean values are reported. The experiments are conducted on a Windows computer equipped with an i7 CPU and 8GB NVIDIA GeForce RTX 2070 GPU. The programming is done by using Keras, a python application program interface (API).

4.1 Dataset description

In this analysis, the Case Western Reserve University (CWRU) rolling bearing dataset is used to test the proposed DA techniques. Acceleration transducers collect data from a single-row deep groove bearing (SKF6205-2RS) at sampling rates of 12 kHz under four load conditions (0 hp, 1 hp, 2 hp, and 3 hp). Depending on the load, this bearing's defect frequencies are multiples of the speed of operation, ranging from 1 797 rpm to 1 730 rpm. The vibration signals used in this analysis are obtained from the motor's drive end under normal (N), inner race fault (OF), outer race fault (IF), and ball fault (BF) health conditions. Single point faults with diameters of 0.18 mm, 0.36 mm and 0.54 mm, respectively, are inserted into the test bearing. As a result, this dataset includes ten bearing health conditions for each of the four loads.

Eighty percent of each set of the original data labeled A is used in training, and 20% is split equally into validation and test sets. The six different artificially created training samples are labeled as follows: 25% overlap (dataset B), 75% overlap (dataset C), Gaussian (dataset D), Kaiser (dataset E), Rectangular (dataset F), and NMF (dataset G). The best-trained model is saved during training, and an early stopping condition truncates the training process when there is no significant improvement in validation loss after ten subsequent epochs.

Table 1 shows the distribution of the original and generated data, and subsequent experimentation substantiates the individual and combinatorial effect on the model's performance.

Table 1 Bearing operating condition training data distribution

Bearing operating condition	Diameter of faults/mm	Size of data								Class label
		Original		Overlap		Window		NMF		
		L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	
Normal	0	480	960	960	1 920	1 440	2 880	480	960	1
Ball	0.18	240	240	480	480	720	720	240	240	2
Ball	0.36	240	240	480	480	720	720	240	240	3
Ball	0.54	240	240	480	480	720	720	240	240	4

Continued

Bearing operating condition	Diameter of faults/mm	Size of data								Class label
		Original		Overlap		Window		NMF		
		L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	L_0	$L_{1,2,3}$	
Inner race	0.18	240	240	480	480	720	720	240	240	5
Inner race	0.36	240	240	480	480	720	720	240	240	6
Inner race	0.54	240	240	480	480	720	720	240	240	7
Outer race	0.18	240	240	480	480	720	720	240	240	8
Outer race	0.36	240	240	480	480	720	720	240	240	9
Outer race	0.54	240	240	480	480	720	720	240	240	10

4.2 Performance evaluation of classification model

The proposed approach is evaluated using a combination of the real and six synthetically generated dataset resulting in 22 experiments for each load condition as shown in Fig. 6.

A	AB	AC	AD	AE	AF	AG	ABC	ABD	ABE	ABF	ABG	ABCD	ABCE	ABCF	ABCG	ABCDE	ABCDF	ABCDG	ABCDEF	ABCDEG	ABCDEFG
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Fig. 6 A depiction of the various dataset combination and annotation

Four performance measures are used to evaluate the different fault classification results. These are accuracy, precision, recall, and f1-score. These can be formulated and computed for a multi-class performance evaluation [25] as

$$\begin{cases} \text{Accuracy} = \dfrac{1}{N_c}\displaystyle\sum_{n=1}^{N_c}\frac{\text{TP}_n+\text{TN}_n}{\text{TP}_n+\text{FP}_n+\text{TN}_n+\text{FN}_n}\times 100 \\ \text{Precision} = \dfrac{1}{N_c}\displaystyle\sum_{n=1}^{N_c}\frac{\text{TP}_n}{\text{TP}_n+\text{FP}_n}\times 100 \\ \text{Recall} = \dfrac{1}{N_c}\displaystyle\sum_{n=1}^{N_c}\frac{\text{TP}_n}{\text{TP}_n+\text{FN}_n}\times 100 \\ F1 = 2\left(\dfrac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\right)\times 100 \end{cases} \tag{5}$$

where N_c is the number of classes, TP and TN are the number of true positive and negative classifications respectively. FP and FN are the number of false positive and negative classifications, respectively.

These metrics directly characterize the requirements of a classification model. False alarms triggered inadvertently are undesirable because they increase operating costs by causing unnecessary downtime. A good classifier will maximize performance by only activating an alert in the presence of an actual fault, ensuring no faults are missed, and no false alarms are generated. The f1-score explicitly expresses this combination.

5. Results and discussion

As mentioned earlier, the time-frequency domain representation of the fault is used as an input sample for training the diagnostic model. The original time-domain representation of the vibration signal is merged with various combinations of augmented samples and utilized as input for the model. The experiment is carried out for different operating speeds and the findings reveal that, the fault classification model performs creditably well in discriminating the various operating states with excellent accuracies as shown in Table 2, where Acc, Pre, Rec, and f1 are the average percentage values of accuracy, precision, recall, and f1-score, respectively. As expected, the computation time increases with an increase in the number of augmentation strategies which is proportional to the size of the training dataset.

Table 2 Experimental results of different combinations of real and augmented datasets %

Dataset	0 hp				1 hp				2 hp				3 hp			
	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1
0	99.85	99.86	99.85	99.85	99.87	99.88	99.87	99.87	99.74	99.75	99.75	99.75	99.62	99.66	99.62	99.62
1	99.98	99.98	99.98	99.98	99.94	99.94	99.94	99.94	99.87	99.88	99.88	99.88	99.81	99.81	99.81	99.81
2	99.92	99.93	99.93	99.93	99.87	99.87	99.87	99.87	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94

Continued

Dataset	0 hp				1 hp				2 hp				3 hp			
	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1	Acc	Pre	Rec	f1
3	99.79	99.79	99.79	99.79	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98
4	99.77	99.78	99.78	99.78	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98
5	99.77	99.78	99.78	99.78	99.62	99.65	99.62	99.62	99.87	99.88	99.88	99.88	99.98	99.98	99.98	99.98
6	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**	**99.98**
7	99.92	99.93	99.93	99.93	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98
8	99.92	99.93	99.93	99.93	99.81	99.81	99.81	99.81	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98
9	99.85	99.86	99.85	99.85	99.87	99.88	99.87	99.87	99.94	99.94	99.94	99.94	99.87	99.88	99.87	99.87
10	99.70	99.71	99.70	99.70	99.81	99.81	99.81	99.81	99.68	99.70	99.69	99.69	99.94	99.94	99.94	99.94
11	99.92	99.93	99.93	99.93	99.87	99.88	99.87	99.87	99.98	99.98	99.98	99.98	99.87	99.88	99.88	99.88
12	99.85	99.86	99.85	99.85	99.81	99.81	99.81	99.81	99.68	99.70	99.70	99.70	99.98	99.98	99.98	99.98
13	99.85	99.86	99.85	99.85	99.98	99.98	99.98	99.98	99.87	99.88	99.87	99.87	99.87	99.88	99.88	99.88
14	99.93	99.93	99.93	99.93	99.98	99.98	99.98	99.98	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98
15	99.92	99.93	99.93	99.93	99.94	99.94	99.94	99.94	99.74	99.78	99.78	99.78	99.98	99.98	99.98	99.98
16	99.85	99.86	99.85	99.85	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98	99.94	99.94	99.94	99.94
17	99.92	99.93	99.93	99.93	99.87	99.88	99.87	99.87	99.81	99.81	99.81	99.81	99.98	99.98	99.98	99.98
18	99.85	99.86	99.85	99.85	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98
19	99.85	99.86	99.85	99.85	99.81	99.82	99.81	99.81	99.87	99.88	99.88	99.88	99.98	99.98	99.98	99.98
20	99.92	99.93	99.93	99.93	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94	99.94
21	99.92	99.93	99.93	99.93	99.94	99.94	99.94	99.94	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98

The computation times for dataset 0 are 141 s, 182 s, 153 s, and 181 s for 0 hp, 1 hp, 2 hp, and 3 hp respectively. It took twice as much time to train and achieve the best-performing model. The computation times for dataset 6 are 345.5 s, 364.5 s, 421.3 s, 389 s for 0 hp, 1 hp, 2 hp, and 3 hp operating speeds, respectively. It took an average of 1 310.25 s to train the largest dataset 21, which is the longest time recorded in training. The average values of the computation times are shown in Fig. 7.

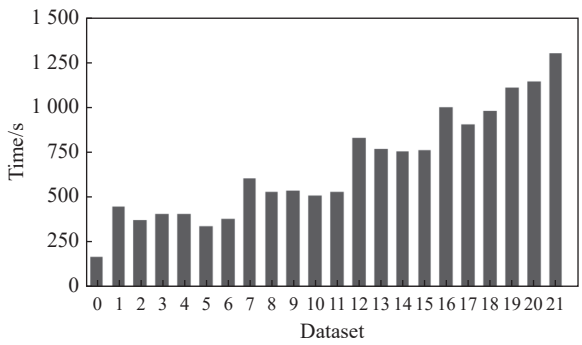


Fig. 7 Average computation time

In subsequent analysis, the model's performance is evaluated by using metrics defined in (5). Fig. 8 shows the test accuracies of the model across all datasets under the four operating speeds.

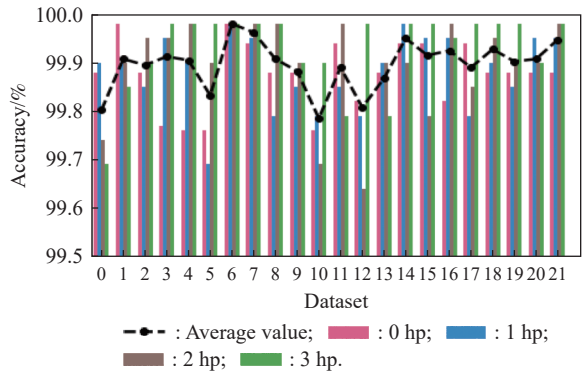


Fig. 8 Test accuracies of all datasets

These accuracies range from 99.78% (dataset 0) to 99.98% (dataset 6). It is worth noting that increasing the dataset size, which necessitates integrating different augmentation strategies, does not always imply better performance. As shown in Fig. 9, dataset 21 has the largest

sample size. However, it has a high computation training time and low accuracy compared to dataset 6, which has a shorter training time and a higher performance outcome.

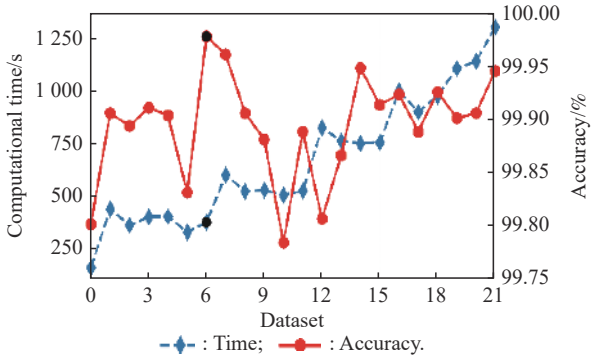


Fig. 9 Computation time against accuracy

Fig. 10 shows the boxplot for all experimental studies with the dataset (0) represented as a baseline and the rest of the datasets split according to the operating conditions.

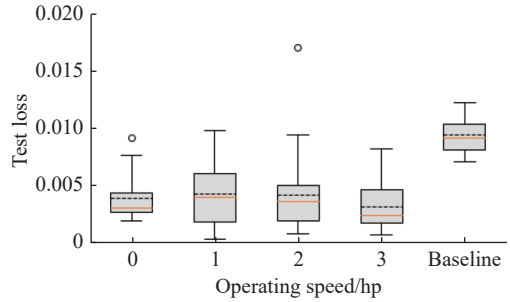


Fig. 10 Test loss for all dataset

In all studies, test loss for all the datasets is compared to that of dataset (0). The average test loss for the baseline is higher than that of the other datasets. Nonetheless, the performance of the proposed methods is excellent across all datasets, and it outperforms most existing fault classification algorithms in many cases.

The findings are compared to recent related works that use the same dataset to demonstrate the proposed method's efficacy and superiority as presented in Table 3.

Table 3 Comparative analysis of the average testing accuracy for different existing models

Method	Description	Number of classes	Training samples	Average test accuracy/%
Method 1 [26]	Ensemble deep auto-encoders	12	2400	97.18
Method 2 [27]	DCNN with frequency domain (FD) features	3	4500	99.38
Method 3 [28]	ELM with spectral Kurtosis	5	–	98.84
Method 4 [29]	Multiscale local feature learning with support vector machine	10	150	99.31
Method 5 [30]	Hierarchical diagnosis network with deep belief network	10	500	99.03
Our method	CNN with augmentation	10	5280 (A + G)	**99.98**

Method 1 is used to detect and classify 12 rolling bearing health conditions by mixing diverse activation functions. The classification accuracy of this approach is 97.18% which is 2.88% lower than that of the proposed method. Methods 2 and 3 consider the diagnosis of fewer health conditions and achieve an accuracy of 99.38% and 98.84%, respectively, which is 0.61% and 1.15% lower than the proposed method. Methods 4 and 5 use fewer labeled samples to diagnose 10 bearing conditions with relatively high accuracy. Even so, the accuracy of the proposed method is 0.67% and 0.96% higher than methods 4 and 5, respectively.

In addition, the cross-domain problem is investigated, in which labeled data at a given operating speed are used for training, and subsequently tested with data on other loads. According to careful analysis of the results obtained, the test accuracy decreases as the value of operating load moves further away from its trained counterpart. The average accuracy obtained by training the model with 0 hp and testing it with 1 hp is 95.85%, decreasing by 7.94% and 15.17% when tested with 2 hp and 3 hp, respectively. As shown in Table 4, this reduction in accuracy value can be observed across all train and test cross-domains.

Table 4 Cross-domain comparative analysis of the average testing accuracy %

Test	Train			
	0 hp	1 hp	2 hp	3 hp
0 hp	–	96.06	90.61	82.27
1 hp	95.83	–	96.54	87.18
2 hp	88.78	99.30	–	98.53
3 hp	83.21	92.69	99.49	–
Average accuracy	89.27	96.02	95.54	89.34

However, training the model with 1 hp and 2 hp yielded very high cross-domain accuracies of 96.02% and 95.54%, respectively. The overall average cross-domain accuracy is 92.54%, as compared to existing methods, with 87.7% [31] and 91.54% [32]. In other existing anal-

yses on cross-domain fault diagnostics, unlabeled testing data is typically assumed to be accessible during the training, making it easier to extract domain-invariable characteristics useful for off-line fault diagnostics. However, in real-world industries applications, testing samples are rarely available ahead of time. The proposed method can learn the significant features of the bearing health conditions from one working domain and generalize the learned pattern to different domains.

6. Conclusions

The acquisition of adequate data in practical applications to enhance existing data on machinery operating conditions is challenging. The quality and volume of available training data play a significant role in the classification accuracy and generalizability of practical DL models on unseen data.

Augmentation of data is an alternate technique to circumvent the unavailability of labeled training data by deliberately synthesizing new labeled samples from currently accessible ones. Conventional DA strategies depend on a minimal collection of established invariances that are simple to invoke and implement offhand with slight disturbances that will not alter the data. In non-image domains, however, such empirical label-preserving transformations are often ineffective.

This paper presents a different approach that utilizes augmented time-frequency signal representations to enhance the training dataset, thereby improving the overall performance of a CNN model. Using the CWRU bearing dataset, the proposed approach is compared to current methods. In order to assess the significance of their technique, the authors conducted many experimental investigations. The findings show that the proposed technique surpasses existing rolling bearing failure classification methods in terms of classification accuracy. In future works, the authors will explore the effect of noise on prediction accuracies. The proposed technique will also be implemented on other forms of time series data used in fault classification problems to substantiate the proposed method further.

References

[1] GOUGAM F, RAHMOUNE C, BENAZZOUZ D, et al. Bearing faults classification under various operation modes using time domain features, singular value decomposition, and fuzzy logic system. Advances in Mechanical Engineering, 2020, 12(10): 1687814020967874.

[2] ROHANI BASTAMI A, BASHARI A. Rolling element bearing diagnosis using spectral kurtosis based on optimized impulse response wavelet. Journal of Vibration and Control, 2020, 26(3/4): 175–185.

[3] ZHANG Y, XING K S, BAI R X, et al. An enhanced convolutional neural network for bearing fault diagnosis based on time-frequency image. Measurement, 2020, 157: 107667.

[4] WANG Z Z, CHEN Y, CAI Z, et al. Methods for predicting the remaining useful life of equipment in consideration of the random failure threshold. Journal of Systems Engineering and Electronics, 2020, 31(2): 415–431.

[5] KULEVOME D K B, WANG H, WANG X G. A bidirectional LSTM-based prognostication of electrolytic capacitor. Progress in Electromagnetics Research C, 2021, 109: 139–152.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012: 1097−1105.

[7] AGBLEY B L Y, LI J, HAQ A, et al. Wavelet-based cough signal decomposition for multimodal classification. Proc. of the 17th International Computer Conference Wavelet Active Media Technology and Information Processing, 2020: 5–9.

[8] AGBLEY B L Y, LI J, HAQ A U, et al. Multimodal melanoma detection with federated learning. Proc of the 18th International Computer Conference Wavelet Active Media Technology and Information Processing, 2021: 238–244.

[9] ZHANG S, ZHANG S B, WANG B N, et al. Deep learning algorithms for bearing fault diagnostics —a comprehensive review. IEEE Access, 2020, 8: 29857–29881.

[10] KULEVOME D K B, WANG H, WANG X G. Deep neural network based classification of rolling element bearings and health degradation through comprehensive vibration signal analysis. Journal of Systems Engineering and Electronics, 2022, 33(1): 233–246.

[11] CHEN X H, ZHANG B K, GAO D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. Journal of Intelligent Manufacturing, 2021, 32(4): 971–987.

[12] KUMAR A, ZHOU Y, GANDHI C P, et al. Bearing defect size assessment using wavelet transform based deep convolutional neural network (DCNN). Alexandria Engineering Journal, 2020, 59(2): 999–1012.

[13] YANG B, LEI Y G, JIA F, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing, 2019, 122: 692–706.

[14] LI X, ZHANG W, DING Q, et al. Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. Journal of Intelligent Manufacturing, 2020, 31(2): 433–452.

[15] LIU S W, JIANG H K, WU Z H, et al. Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. Mechanical Systems and Signal Processing, 2022, 163: 108139.

[16] HU T H, TANG T, LIN R L, et al. A simple data augmentation algorithm and a self-adaptive convolutional architecture for few-shot fault diagnosis under different working conditions. Measurement, 2020, 156: 107539.

[17] GAO X, DENG F, YUE X H. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty. Neurocomputing, 2020, 396: 487–494.

[18] AHMED H, NANDI A K. Condition monitoring with vibration signals: compressive sampling and learning algorithms for rotating machines. New Jersey: John Wiley & Sons, 2020.

[19] IWANA B K, UCHIDA S. An empirical survey of data augmentation for time series classification with neural networks. PLoS One, 2021, 16(7): e0254841.

[20] JIN K H, MCCANN M T, FROUSTEY E, et al. Deep convolutional neural network for inverse problems in imaging. IEEE Trans. on Image Processing, 2017, 26(9): 4509–4522.
[21] NYKAZA E, BUNKLEY S, BLEVINS M G. Objectively choosing spectrogram parameters to classify environmental noises. INTER-NOISE and NOISE-CON Congress and Conference Proceedings, 2006, 253 (1): 7336–7343.
[22] GAN J Z, LIU T, LI L, et al. Non-negative matrix factorization: a survey. The Computer Journal, 2021, 64(7): 1080–1092.
[23] BERRY M W, BROWNE M, LANGVILLE A N, et al. Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics & Data Analysis, 2007, 52(1): 155–173.
[24] COHEN L. Time-frequency analysis. New Jersey: Prentice Hall, 1995.
[25] ALLEN R L, MILLS D. Signal analysis: time, frequency, scale, and structure. New Jersey: John Wiley & Sons, 2004.
[26] SHAO H D, JIANG H K, LIN Y, et al. A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. Mechanical Systems and Signal Processing, 2018, 102: 278–297.
[27] TIAN Y L, LIU X Y. A deep adaptive learning method for rolling bearing fault diagnosis using immunity. Tsinghua Science and Technology, 2019, 24(6): 750–762.
[28] UDMALE S S, SINGH S K. Application of spectral kurtosis and improved extreme learning machine for bearing fault classification. IEEE Trans. on Instrumentation and Measurement, 2019, 68(11): 4222–4233.
[29] LI J M, YAO X F, WANG X D, et al. Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis. Measurement, 2020, 153: 107419.
[30] GAN M, WANG C, ZHU C A. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mechanical Systems and Signal Processing, 2016, 72: 92–104.
[31] XU Z F, LI C, YANG Y. Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks. Applied Soft Computing, 2020, 95: 106515.
[32] LI X, ZHANG W, DING Q. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. Signal Processing, 2019, 161: 136–154.

Biographies

KULEVOME Delanyo Kwame Bensah was born in 1983. He received his M.E. degree in electronic science and engineering in 2019 from the University of Electronic Science and Technology of China, Chengdu, China, where he is currently pursuing his Ph.D. degree in information and communication engineering. His research interests include prognostics and health management of systems, fault diagnostics, signal processing, and deep learning.
E-mail: kdelanyo@ieee.org

WANG Hong was born in 1974. He received his B.S., M.S. and Ph.D. degrees from Northwestern Polytechnical University, Chongqing University, and University of Electronic Science and Technology of China (UESTC), respectively. He is a faculty member with UESTC since 2003. From 2007 to 2009 he was engaged in doctoral research with the Second Research Institute of Civil Aviation Administration. From 2009 to 2010, he was a research scholar with Polytechnic Institute of New York University and research assistant with New Jersey Institute of Technology, USA. His research interests include radar signal processing, avionics, aeronautical telecommunication, and surveillance technologies in air traffic control.
E-mail: hongw@uestc.edu.cn

WANG Xuegang was born in 1962. He received his Ph.D. degree from Xidian University in 1992. He is now a professor and Ph.D. supervisor with University of Electronic Science and Technology of China. His research interests include radar signal processing, and millimeter wave radar.
E-mail: xgwang@uestc.edu.cn