Novel Convolutional Neural Network (NCNN) for the Diagnosis of Bearing Defects in Rotary Machinery

Anil Kumar, Govind Vashishtha, C. P. Gandhi, Yuqing Zhou, Adam Glowacz, and Jiawei Xiang, *Member, IEEE*

***Abstract*— This work presents the development of novel convolutional neural network (NCNN) for effective identification of bearing defects from small samples. For effective feature learning from small training data, cost function of convolution neural network (CNN) is modified by adding additional sparsity cost in the existing cost function. A novel trigonometric cross-entropy function is developed to compute the sparsity cost. The proposed cost function introduces sparsity by avoiding unnecessary activation of neurons in the hidden layers of CNN. For identification of bearing defects from small training samples, NCNN-based transfer learning is applied in the following manner. First, the raw vibration signals as well as envelope signals from source domain machine are obtained. Thereafter, these envelope signals are applied to NCNN for the learning of features from the big training data acquitted from the source domain. After feature learning, knowledge gained from NCNN is transferred to do fine-tuning of NCNN from small training samples of target domain. Thereafter, defect identification is carried out by applying the test data of target domain to fine-tuned NCNN. The experimental result validates that the proposed cross-entropy function introduces sparsity in CNN and, hence, creates an effective deep learning which can even work under a situation when training data are not available in abundant.**

***Index Terms*— Cost function, intelligent condition monitoring, novel convolutional neural network (NCNN), online diagnostic, transfer learning, trigonometric cross-entropy function.**

Manuscript received September 30, 2020; revised January 7, 2021; accepted January 16, 2021. Date of publication February 1, 2021; date of current version February 17, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1909217 and Grant U1709208, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LD21E050001, and in part by the Zhejiang Special Support Program for High-Level Personnel Recruitment of China under Grant 2018R52034. The Associate Editor coordinating the review process for this article was Dr. Loredana Cristaldi. *(Corresponding author: Jiawei Xiang.)*

Anil Kumar is with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China, and also with Amity University Uttar Pradesh, Noida 201313, India (e-mail: anil_taneja86@yahoo.com).

Govind Vashishtha is with the Sant Longowal Institute of Engineering and Technology, Longowal 148106, India (e-mail: govindyudivashishtha@gmail.com).

C. P. Gandhi is with the Department of Mathematics, Rayat Bahra University, Mohali 140104, India (e-mail: cchanderr@gmail.com).

Yuqing Zhou and Jiawei Xiang are with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China (e-mail: zhouyq@wzu.edu.cn; wxw8627@163.com).

Adam Glowacz is with the Department of Automatic Control and Robotics, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, 30059 Cracow, Poland (e-mail: adglow@agh.edu.pl).

Digital Object Identifier 10.1109/TIM.2021.3055802

I. Introduction

MACHINES are the core element in any manufacturing unit in the era of Industry 4.0. Any fault within the system not only increases the risk of failure but also leaves an impact on machine yield, production quality, and maintenance cost [1].

Bearings and gears are widely used as important machine components but most prone to defects. Their improper functioning leads to catastrophic failure of the machinery. The performance of bearing and gear directly or indirectly influences the operational reliability of the machines [2], [3]. To enhance the reliability of bearings and gears, there is an urgent demand for an effective fault diagnosis technique. Several researchers developed different fault diagnosis methods for identifying defects in bearing and gear. The conventional defect diagnosis methods are generally hinged on frequency domain, time domain, and time–frequency domain techniques [4], [5]. The conventional data-driven fault diagnosis approach, based on statistical features, is found to be efficacious for classifying different health states. In order to reckon the most reliable fault features under harsh environment, removal of noise is essential. That is why denoising of signals have attracted many researchers [2], [6]–[8]. The classification and learning models, such as support vector machines (SVMs) and artificial neural network (ANN), are found to be beneficial for intelligent fault diagnosis [9]–[11]. However, these models completely rely on manual feature engineering [12].

To overcome the problematic situation under manual feature learning, researchers have evolved many advanced fault diagnosis techniques. Deep learning is gaining popularity because of its capability for automatically extracting fault features and valuable information from the collected data and, hence, proved to be more advantageous than any traditional method. Among various deep learning methods, the convolution neural network (CNN) has been successfully applied in various fields like image processing, speech recognition, signal processing, and so on [13]. CNN takes raw data and is capable of performing various operations such as convolution, nonlinear activation, and pooling for extracting high-level features. In comparison with the traditional methods, CNN is capable of analyzing and extracting more robust features which have better effectiveness [14]. Moreover, CNN reduces computation timing and training cost under weight sharing and pooling

operations. Thus, CNN can be considered as an important tool for the identification of defects in bearing and gear. Guo *et al.* [15] introduced hierarchical adaptive CNN for pattern recognition and evaluation of fault size in bearing. Azamfar *et al.* [16] carried out fusion of multisensor current data for carrying out diagnosis of gear box defects. When compared with machine learning techniques based on hand-crafted features, the authors proved the effectiveness of their proposed approach over conventional methods. Chen *et al.* [17] applied CNN for identifying gear defects and established that CNN can provide better classification efficiency in comparison with SVM.

CNN has very good recognition accuracy under persistent operating conditions. However, in real-world applications, machinery is operated under varying load and speed conditions. This reduces the performance of CNN more rapidly. This issue of CNN can be, however, resolved with the help of a new branch of deep learning names as transfer learning which uses a knowledge-based technique to develop a model for little target which may or may not have the label [18]. In transfer learning, the probability distribution functions for training and testing samples are usually different. It is known as a domain adaptation branch and can be applied to various domains for achieving effective results [19]–[21]. Transfer learning is a powerful deep leaning method in which preprocessing is not necessary. It can be classified into various categories including instance learning, feature transfer learning, parameter learning, and relational knowledge transfer learning [19]. Feature transfer learning can represent the target domain into "good" and encode the knowledge into a learned feature. Transfer learning is gaining importance in various fields, such as image recognition, character recognition, emotion analysis, and so on [22]–[25]. Some scholars have also applied transfer learning in mechanical fault diagnosis. Tang *et al.* [26] introduced an effective fault diagnosis method which can enhance the classification accuracy under different working conditions. Wen *et al.* [25] applied automatic encoder for the purpose of identifying bearing defect. Chen *et al.* [27] extracted the cross-domain features by utilizing the transfer components for the diagnosis of faults. The cross-domain approach is adopted for obtaining a space in which the domains are close to each other. Knowledge is transferred from one domain to another domain. However, when the target machine training samples are very small and the classification size is large, the transfer learning does not perform well. This happens because the convolution layer of CNN has many parameters which results in overfitting of the model. Overfitting of the model can be avoided by reducing the activation of neurons which can be achieved by introducing a sparsity cost in the existing function of CNN. Subsequently, a novel trigonometric cross-entropy function is developed in this study to compute the sparsity cost. Initially, the desired sparsity or average activation is specified. The proposed cross-entropy function calculates divergence from sparsity. If there is a divergence from desired sparsity, the penalty of sparsity will be added to existing function. This would affect the activation of hidden layers and eventually result in improvement of learning.

The proposed novel convolutional neural network (NCNN)-based transfer learning methodology is as follows. First, broad sample training data from the source domain are feeded to the improved CNN for learning of defect features. After successful learning, features learned by a few layers of NCNN are transferred for learning of fault knowledge from small training samples. The learning from small samples is called fine-tuning using data of the target machine. Thereafter, the knowledge gained from NCNN is transferred to do fine-tuning of NCNN from small training samples. The test data of target domain are then applied to fine-tuned NCNN for the purpose of identification of defects. The proposed cross-entropy function introduces sparsity in CNN and, hence, creates an effective deep learning which can work under a situation when training data are not available in abundant.

The rest of the proposed work is arranged as follows. Section II introduces the underlying theory and mathematical concepts of CNN. Section III provides the proposed transfer learning-based methodology for the identification of bearing faults. The experimental study and applicability of the proposed methodology is provided in Section IV, whereas Section V provides the concrete conclusions and future scope of the present work.

II. Underlying Theory

A. Convolutional Neural Network

The operation of various CNN layers can be mathematically described [28]–[32] as follows.

Convolution Layer: The convolution operation for each layer of CNN can be mathematically expressed as follows [33]:

$$Cl_f^s(y) = \varphi\left[\left(\sum_{ch=1}^{CH}\sum_{k=1,x=p}^{k=t,x=p+t}\right)\left(W_f^{\text{con}^s}(k)\times Cl_c^{s-1}(x)\right)+b_f^{\text{con}}\right]. \tag{1}$$

In (1), the pixel value for sth layer of fth filter at yth position is represented by $Cl_f^s(y)$. Similarly, for the chth channel, the pixel value of $s-1$th convolutional layer at position x is given by $Cl_c^{s-1}(x)$, where s and CH are initial pixel location and the total number of channels, respectively. $W_f^{\text{con}^s}(k)$ indicates weight of the sth layer at kth position and b_f^{con} signifies bias term of the fth filter. The total element for the same filter is represented by t.

Three convolution layers and sigmoid transfer function (φ) have been used to model CNN.

Maximum Pooling Layer: The pooling operation of CNN can be mathematically expressed as follows

$$M_c^y(y) = \max\left(L_f^s(x)\right) \text{ for } x = 1, \quad 1 \text{ to } \text{pat}_h, \text{pat}_w. \tag{2}$$

In (2), $M_c^y(y)$ is the pixel value obtained after application of maximum pooling on sth layer of chth channel having patch height pat_h and image width of pat_w, respectively.

Fully connected layer: (3) shows the mathematical procedure adopted by a fully connected layer where $f\text{eat}_k$ represents kth input feature vector

$$I_{\text{fc}} = \varphi\left[\sum_{k=1}^{K}\left(f\text{eat}_k \times W_{k,j}^{\text{fc}}\right) + b_j^{\text{fc}}\right]. \tag{3}$$

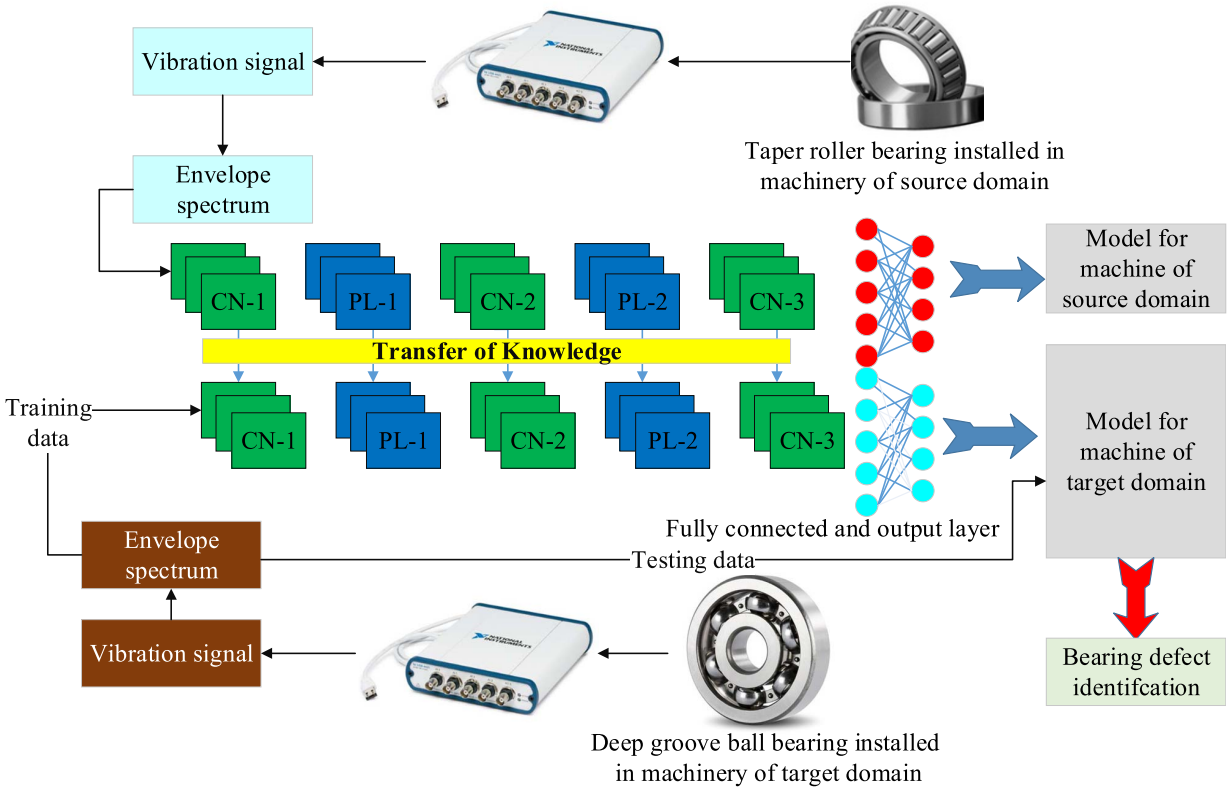


Fig. 1. Schematic flowchart of the proposed transfer learning-based methodology.

The weight of kth input feature of jth hidden layer neuron is $W_{k,j}^{\text{fc}}$ having added bias of b_j^{fc}. The term I_{fc} gives the output for jth hidden layer neuron and K is total input features.

SoftMax layer: The mathematical representation of SoftMax layer, which predicts the output of fault conditions (fc), is given by (4). In this study, 12 different fault conditions are analyzed by the proposed methodology.

This layer computes the loss experienced during the training phase. The existing cost function (5) is an objective function which must be minimized in order to predict the data. CNN minimizes the loss calculated by a SoftMax layer

$$P_{\text{so}} = \frac{\exp(I_{\text{fc}})}{\sum_{fc=1}^{\text{fc}} \exp(I_{\text{fc}})}. \tag{4}$$

The existing cost function is given as

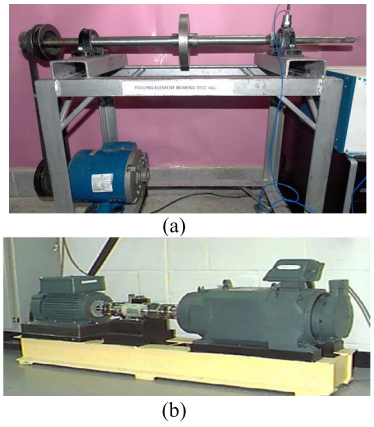
$$e^{\text{modified}} = \text{CE} + \beta \sum w^2. \tag{5}$$

The fully connected layer has many parameters that adhere to deep learning from small data. In this work, a sparsity cost is added in the existing cost function (5) which can make the learning deep. The sparsity is incorporated by avoiding unnecessary activation of neurons in the hidden layer of CNN

$$e^{\text{modiifed}} = \text{CE} + \beta \sum w^2 + \beta 1 \sum_{j=1}^{M} N\left(\mu, \mu_j^T\right) \tag{6}$$

where

$$\text{CE} = \sum_{j=1}^{m} y_j^T \ln y_j^P \tag{7}$$



(a)

(b)

Fig. 2. Source and target domain machine. (a) Source domain. (b) Target domain.

and

$$N\left(\mu, \mu_j^T\right) = \sum_{j=1}^{J} \left[\begin{array}{l} -6\tan 1 + \left(2+\mu+\mu_j^T\right) \tan\left[\dfrac{8+\left(\mu^2+\left(\mu_j^T\right)^2\right)\left(\mu+\mu_j^T\right)^2}{8\left(1+\mu^2\left(\mu_j^T\right)^2\right)}\right] \\ +\left(4-\mu-\mu_j^T\right) \tan\left[\dfrac{8+\left((1-\mu)^2+\left(1-\mu_j^T\right)^2\right)\left(2-\mu-\mu_j^T\right)^2}{8\left(1+(1-\mu)^2\left(1-\mu_j^T\right)^2\right)}\right] \end{array} \right]. \tag{8}$$

Here, y^P represents the predicted value, y^T represents the target value, m is training data, the regularization factor (L_2)

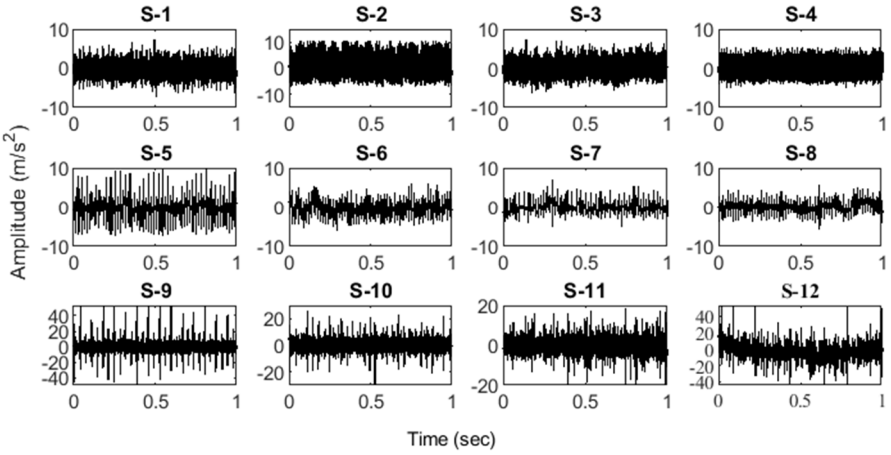


Fig. 3. Time-domain vibration data of source domain machine under different health conditions.

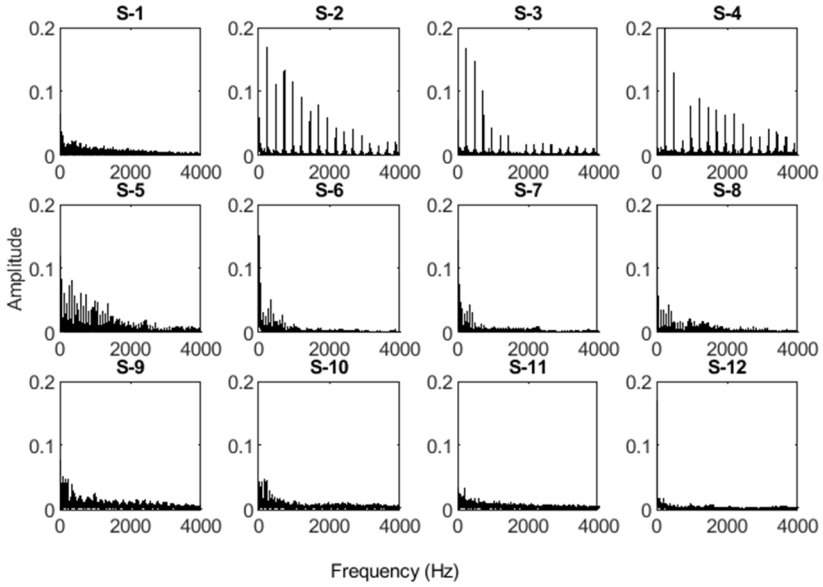


Fig. 4. Frequency-domain vibration data of source domain machine under different health conditions.

is denoted by β, $\beta 1$ is the penalty of sparsity and j indicates total neurons in hidden layer.

III. Methodology

The schematic flowchart of the proposed transfer learning-based fault diagnosis methodology is depicted in Fig. 1 and works as follows.

Step 1: Time-domain vibration data are acquired from the source domain machine.

Step 2: Envelope spectrum of vibration data is computed.

Step 3: Modification of existing cost function of CNN is done to form NCNN.

Step 4: Training of NCNN using training data (envelope spectrum) of the source domain.

Step 5: Transfer of knowledge by transferring weight and bias parameters of hidden layers. The layer whose parameters are transferred can be seen in Fig. 1.

Step 6: Obtain target domain data sets. Thereafter, fine-tuning of NCNN using the training data set of the target machine is done by minimizing the modified cost function of NCNN till appropriate performance is achieved.

Step 7: Apply the test data of the target machine to fine-tuned NCNN for the identification of defects.

IV. Experiment and Application of Proposed Method

The efficiency of proposed scheme is validated on the bearing data set of two different domains, one is source domain, and another is called target domain. The source machine [shown in Fig. 2(a)] consists of 0.5 HP motor (left), from which power is transmitted to the shaft using pulley arrangement. The motor shaft is supported by test bearings. Single groove was introduced to the test bearings using electro-discharge machining. The detail of defect conditions

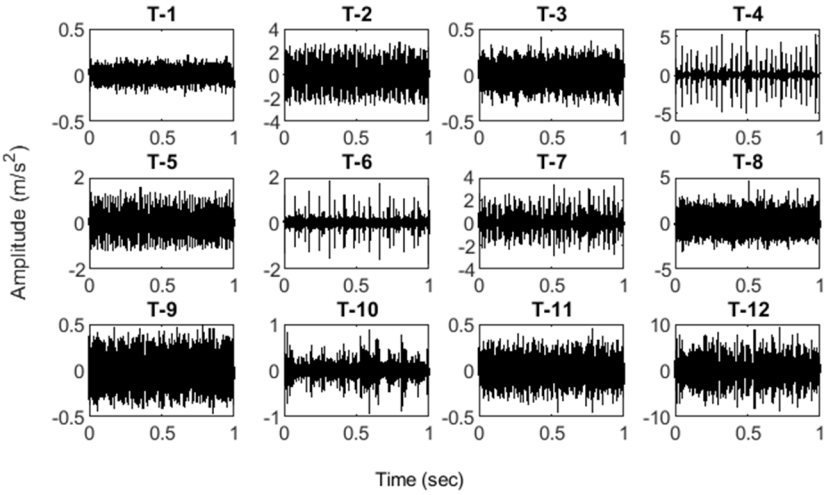


Fig. 5. Time-domain vibration data of target domain machine under different health conditions.

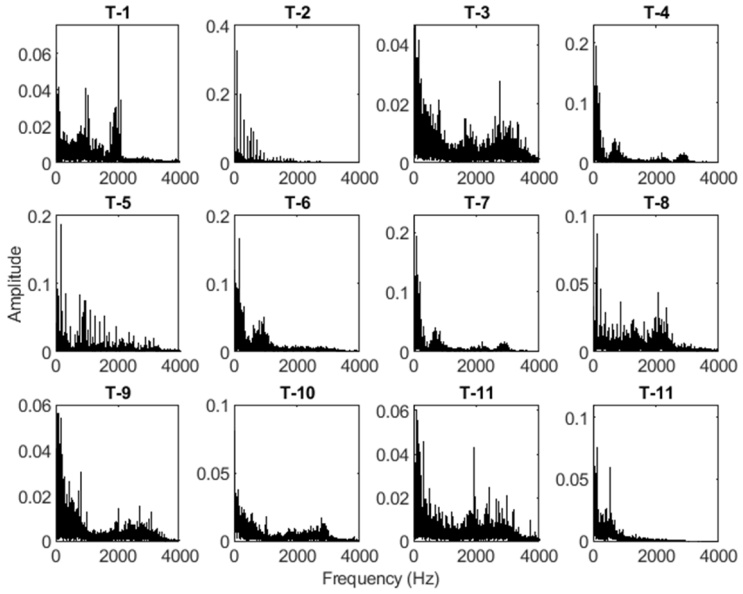


Fig. 6. Frequency-domain vibration data of the target domain machine under different health conditions.

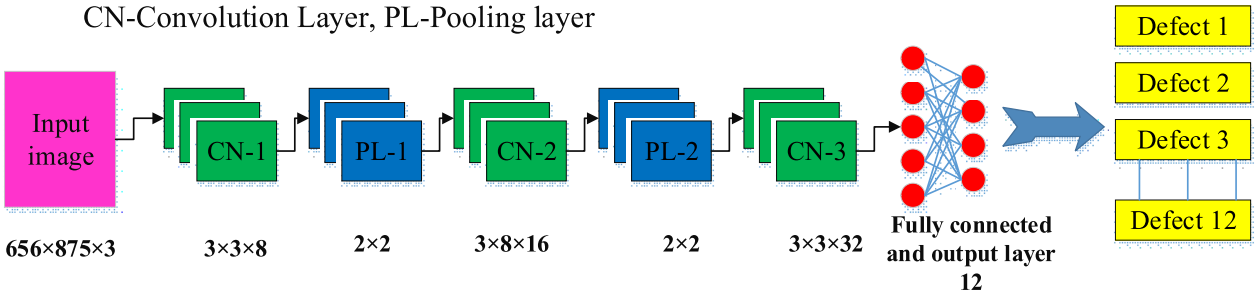
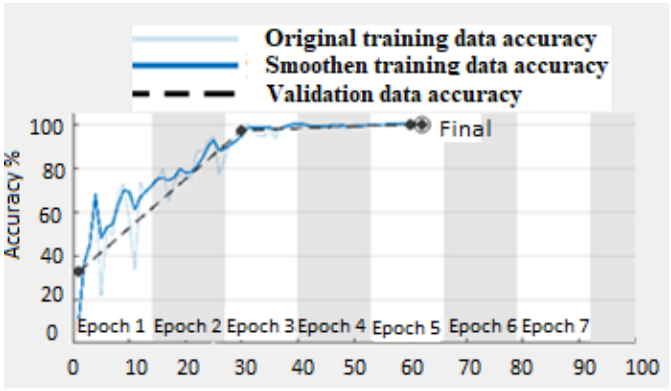


Fig. 7. Structure of NCNN.

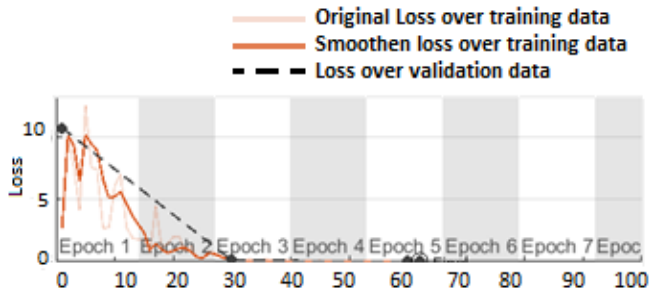
under study is given in Table I. For further detail of test rig, please refer to articles [7], [34]. Vibration data were collected using an accelerometer, which was attached to the housing. Vibration data were acquired at the rate of 70 000 samples per second. The vibration signals obtained from source domain machine having defect conditions, mentioned in Table I, are shown in Fig. 3. For converting time-domain signals into frequency domain, envelope spectrums were computed and

TABLE I
DETAIL OF SOURCE AND TARGET DOMAIN TRAINING AND TESTING DATA

Domain	Defect Condition	Label	Speed/ Sampling Frequency /Load	Training Sample	Testing sample
Source	Normal Condition	S-1	2050 rpm/ 70000 Sample/sec/0.16 HP	2400 (200 ×12)	240 (20× 12)
	Outrace defect of 22.4 Mils	S-2			
	Outrace defect of 46.4 Mils	S-3			
	Outrace defect of 67.7 Mils	S-4			
	Inner race defect of 18.5 Mils	S-5			
	Inner race defect of 40.5 Mils	S-6			
	Inner race defect of 58.6 Mils	S-7			
	Inner race defect of 71.2 Mils	S-8			
	Ball defect of 18.1 Mils	S-9			
	Ball defect of 44.0 Mils	S-10			
	Ball defect of 56.6 Mils	S-11			
	Ball defect of 79.1 Mils	S-12			
Target	Normal Condition	T-1	1750 rpm/ 12000 Sample/sec/20 HP	240 (20 ×12)	360 (30× 12)
	Outrace defect of 7 Mils	T-2			
	Outrace defect of 14	T-3			
	Outrace defect of 21	T-4			
	Inner race defect of 7	T-5			
	Inner race defect of 14	T-6			
	Inner race defect of 21 Mils	T-7			
	Inner race defect of 28	T-8			
	Ball defect of 7 Mils	T-9			
	Ball defect of 14 Mils	T-10			
	Ball defect of 21 Mils	T-11			
	Ball defect of 28 Mils	T-12			



(a)



(b)

Fig. 8. Training performance achieved with source data. (a) Training performance. (b) Cost function.

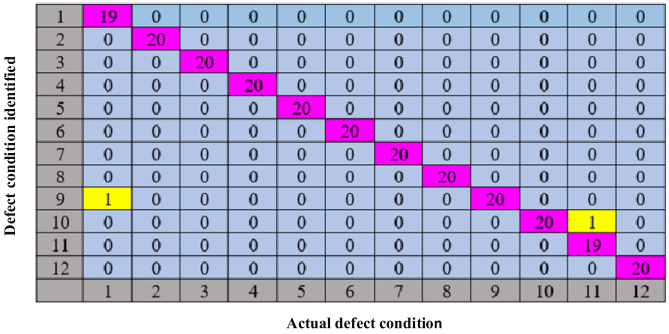
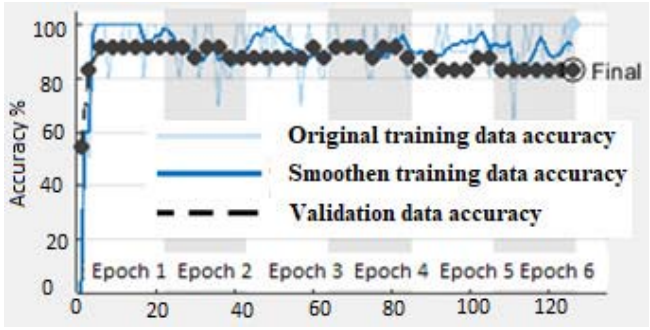


Fig. 9. Confusion matric of result achieved with source data.

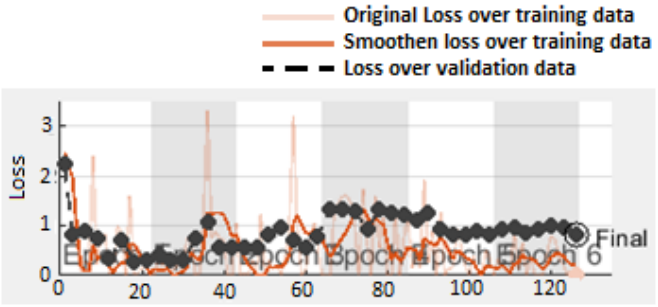
are shown in Fig. 4. Each envelope spectrum was saved as jpeg RGB image of size 656 × 875 × 3 for further feeding into CNN. The 656 indicates width, 875 indicates length, and three indicates dimensionality of image. The software platform used for processing the data and developing the model was MATLAB 2019a with deep learning toolbox.

From the source machine, 2400 samples from a total of 12 conditions with 200 samples in each condition were used to train the improved CNN. The target machine is a test rig of Case Western Reserve University [35] and is shown in [Fig. 2(b)]. The data from the target machine were acquired with the sampling rate of 12 000 samples per second. The vibration signals obtained from the target domain machine having defect conditions mentioned in Table I are shown in Fig. 5. The corresponding frequency domain signals are shown in Fig. 6.

First, data from the source domain are applied to NCNN, and the structure of CNN used in this work is shown in Fig. 7. The training performance and loss function of NCNN while training with source domain data are shown in Fig. 8(a) and (b), respectively. After training, test data of the source domain are applied to NCNN. The results are shown in Fig. 9. After this, the knowledge gained over source domain data is transferred and is utilized to learn defect features from target domain data. Then, training data (envelope spectrum) of target domain data (240 samples, 20 samples of each condition) are applied to NCNN for the fine-tuning of NCNN. These 240 samples of target domain data were obtained from a total of 12 conditions mentioned in Table I. The model has been developed specially to learn effectively from small training samples. This is the reason why Target machine has been developed from very few samples (240 samples, 20 of each defect condition). The training performance and loss function of NCNN over target domain data are shown in Fig. 10(a) and (b), respectively. After training, test data of



(a)



(b)

Fig. 10. Training performance achieved with target data. (a) Training performance. (b) Cost function.

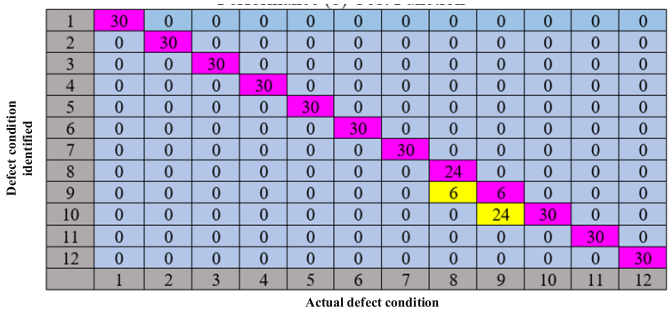


Fig. 11. Confusion matrix of the result achieved with target data (with transfer learning).

target domain (360 samples, 30 samples of each condition) are applied to fine-tuned NCNN. The defect identification results of target domain data are presented in Fig. 11. The overall accuracy of the proposed scheme is found to be 91%, which is shown in Table II.

A. Comparison With Other Different Methods

To prove the efficacy of the proposed transfer learning strategy, a comparison has also been made by applying different transfer strategies such as feature transfer learning, domain adaptation [36], and joint distribution learning [37]. The defect identification results are presented in Table II. The modified cost function along with the sigmoid transfer function gives defect identification accuracy of 91%. The defect identification accuracy without transfer learning is very less (highest accuracy achieved is 51% with Sigmoid + modified cost function) in comparison with that achieved using

TABLE II
COMPARISON WITH DIFFERENT METHODS

Transfer strategy		Accuracy
RELU +Existing cost function	Transfer learning	319/360(0.88)
Leaky RELU+ Existing cost function		316/360(0.87)
Clipped RELU +Existing cost function		315/360(0.87)
Exponential linear unit+ Existing cost function		318/360(0.88)
Tanh + Existing cost function		319/360(0.88)
Sigmoid+ Existing Cost Function		319/360(0.88)
Sigmoid + Modified cost function (proposed method)		330/360(0.91)
Domain adaptation (DAN)[36]		323/360 (0.89)
Joint distribution learning [37]		320/360(0.88)
RELU +Existing cost function	Without transfer learning	172/360(0.47)
Leaky RELU+ Existing cost function		183/360(0.50)
Clipped RELU +Existing cost function		176/360(0.48)
Exponential linear unit+ Existing cost function		179/360(0.49)
Tanh + Existing cost function		178/360(0.49)
Sigmoid+Existing cost function		180/360(0.50)
Sigmoid + Modified cost function		185/360(0.51)

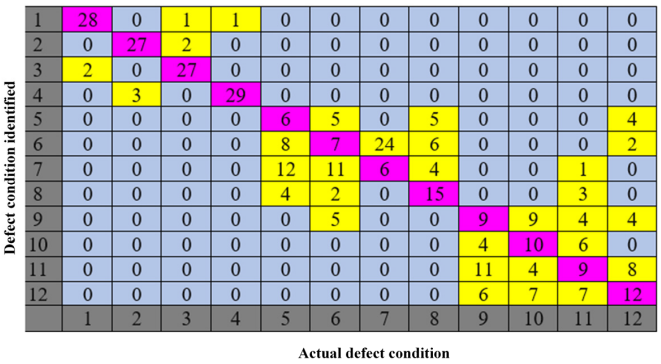


Fig. 12. Confusion matrix of the result achieved with target data (without transfer learning).

transfer learning. The confusion matrix of result obtained without transfer learning is presented in Fig. 12. The defect identification accuracy in case of inner race (5–8) and rolling element defect (9–12) is very low. It is because of the fact that defect on the moving components (inner race and outer race) results in highly nonstationary signature. As a result, training with small data set of such defects results in under learning because network learns from scratch. On the other hand, using transfer learning, the pregained knowledge attained over source domain data helps CNN in learning target domain defects conditions from small data set. A comparative analysis of the results, presented in Table II, shows that the identification of defects through the proposed NCNN-based transfer learning is much better than the existing CNN-based

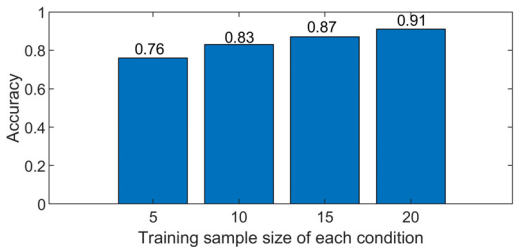


Fig. 13. Effect of variation of training sample size on the accuracy.

transfer learning, when training data size is not available in abundant.

B. Effect of Training Sample Size on the Network Performance

A study has also been made to test the accuracy of the proposed model by varying the training sample size. The model has been effectively developed especially to learn from small training samples. Therefore, the identification accuracy of the proposed model has been tested by decreasing the sample size. The results are presented in Fig. 13. The accuracy is found to be decreased with the decrease in training sample size.

V. Conclusion

An improved transfer learning technique is introduced for the identification of defects in bearing from a small training data set. A modification is proposed in the existing cost function of CNN, which is the novelty of this research. The modification evades the redundant activation of neurons in the hidden layers of CNN. For computing sparsity cost, a trigonometric cross-entropy function is developed, which is further added in the existing cost function of CNN which calculates divergence from desired sparsity. If there is a divergence, the penalty of sparsity has been added to existing function. This has affected the activation of hidden layers and eventually resulted in improvement of learning. The improved CNN performs much better in comparison with the existing methods based on CNN. The combination of modified cost function along with the sigmoid transfer function gives 91% defect identification accuracy. The accuracy without transfer learning is significantly less (highest accuracy achieved is 51% with sigmoid + modified cost function) in comparison with that achieved using the proposed transfer learning-based methodology. The experiment result suggests that the proposed transfer strategy can be deployed to transfer knowledge gained by the network in the environment where training data are available in abundant. This can also be used to develop fined-tuned model under the situation where training data are not abundant. In future work, authors wish to develop numerical simulation compliance artificial intelligence models which can be included in portable devices using field-programmable gate arrays (FPGAs) for bearing defects identification.

Appendix

The proposed trigonometric fuzzy cross-entropy loss function, which can measure divergence between the desired activation value μ and average activation value μ_J^T of the hidden layer J over the training set, can be established as follows.

Defination 1: [38] A function $N\left(\mu, \mu_j^T\right)$, represented by (8), is a valid fuzzy cross-entropy loss function iff it satisfies the following essential properties.

$(i) N\left(\mu, \mu_j^T\right) \geq 0$ with equality if $\mu = \mu_j^T (ii) N\left(\mu, \mu_j^T\right)$ is a convex function of both μ and μ_j^T where $\mu, \mu_j^T \in [0, 1]$.

Proof: We shall first establish the following lemma, intended to establish the nonnegativity of $N\left(\mu, \mu_j^T\right)$ as follows.

Lemma 1: For $\mu, \mu_j^T \in [0, 1]$, there exists the inequality

$$\sqrt{\frac{\mu^2 + \left(\mu_j^T\right)^2}{2}} \geq \frac{2\mu\mu_j^T}{\mu + \mu_j^T}. \tag{A.1}$$

Proof: If $S = \sqrt{(\mu^2 + \left(\mu_j^T\right)^2/2)}$ and $H = (2\mu\mu_j^T/\mu + \mu_j^T)$, then the resulting inequality (A.1) shall be true if

$$\begin{aligned} S^2 \geq H^2 \Rightarrow \frac{\mu^2 + \left(\mu_j^T\right)^2}{2} &\geq \frac{4\mu^2\left(\mu_j^T\right)^2}{\left(\mu + \mu_j^T\right)^2} \\ \Rightarrow \left(\mu^2 + \left(\mu_j^T\right)^2\right)\left(\mu^2 + \left(\mu_j^T\right)^2 + 2\mu\mu_j^T\right) &\geq 8\mu^2\left(\mu_j^T\right)^2 \\ \Rightarrow \left(\mu^2 + \left(\mu_j^T\right)^2\right)^2 + 2\mu\mu_j^T\left(\mu^2 + \left(\mu_j^T\right)^2\right) & \\ + \mu^2\left(\mu_j^T\right)^2 \geq 9\mu^2\left(\mu_j^T\right)^2 & \\ \Rightarrow \left(\mu^2 + \left(\mu_j^T\right)^2 + \mu\mu_j^T\right)^2 - 9\mu^2\left(\mu_j^T\right)^2 &\geq 0 \\ \Rightarrow \left(\mu - \mu_j^T\right)^2\left(\mu^2 + \left(\mu_j^T\right)^2 + 4\mu\mu_j^T\right) &\geq 0 \end{aligned} \tag{A.2}$$

which is obviously true for each $\mu, \mu_j^T \in [0, 1]$ with equality if $\mu = \mu_j^T$.

Thus, in view of resulting Lemma 1, we have

$$\begin{aligned} \sqrt{\frac{\mu^2 + \left(\mu_j^T\right)^2}{2}} &\geq \frac{2\mu\mu_j^T}{\mu + \mu_j^T} \\ \Rightarrow \left(\mu^2 + \left(\mu_j^T\right)^2\right)\left(\mu + \mu_j^T\right)^2 &\geq 8\mu^2\left(\mu_j^T\right)^2. \end{aligned} \tag{A.3}$$

$$\begin{aligned} \Rightarrow \frac{\left(\mu^2 + \left(\mu_j^T\right)^2\right)\left(\mu + \mu_j^T\right)^2}{8} + 1 &\geq \mu^2\left(\mu_j^T\right)^2 + 1 \\ \Rightarrow \frac{8 + \left(\mu^2 + \left(\mu_j^T\right)^2\right)\left(\mu + \mu_j^T\right)^2}{8\left(1 + \mu^2\left(\mu_j^T\right)^2\right)} &\geq 1. \end{aligned} \tag{A.4}$$

Using the monotonicity property of tangent function over [0, 1], the foregoing inequality yields

$$\begin{aligned} \left(2 + \mu + \mu_j^T\right)\tan\left(\frac{8 + \left(\mu^2 + \left(\mu_j^T\right)^2\right)\left(\mu + \mu_j^T\right)^2}{8\left(1 + \mu^2\left(\mu_j^T\right)^2\right)}\right) & \\ \geq \left(2 + \mu + \mu_j^T\right)\tan 1. \end{aligned} \tag{A.5}$$

Interchanging μ, μ_j^T with their counterparts, the forthcoming inequality (A.5) yields

$$\left(4-\mu-\mu_j^T\right)\tan\left(\frac{8+\left((1-\mu)^2+\left(1-\mu_j^T\right)^2\right)\left(2-\mu-\mu_j^T\right)^2}{8\left(1+(1-\mu)^2\left(1-\mu_j^T\right)^2\right)}\right) \geq \left(4-\mu-\mu_j^T\right)\tan 1. \quad \text{(A.6)}$$

Simply adding the resulting inequalities (A.5) and (A.6) and taking the summation over $j = 1$ to yields $N\left(\mu, \mu_j^T\right) \geq 0 \forall \mu, \mu_j^T \in [0,1]$ with equality if $\mu = \mu_j^T$. Furthermore, the function $N\left(\mu, \mu_j^T\right)$ admits the convexity property with respect to each μ, μ_j^T as can be seen from Mathematica software.

REFERENCES

[1] L. Xiao, J. Tang, X. Zhang, and T. Xia, "Weak fault detection in rotating machineries by using vibrational resonance and coupled varying-stable nonlinear systems," *J. Sound Vib.*, vol. 478, Jul. 2020, Art. no. 115355, doi: 10.1016/j.jsv.2020.115355.

[2] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Variational mode decomposition based symmetric single valued neutrosophic cross entropy measure for the identification of bearing defects in a centrifugal pump," *Appl. Acoust.*, vol. 165, Aug. 2020, Art. no. 107294, doi: 10.1016/j.apacoust.2020.107294.

[3] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Latest developments in gear defect diagnosis and prognosis: A review," *Measurement*, vol. 158, Jul. 2020, Art. no. 107735, doi: 10.1016/j.measurement.2020.107735.

[4] S. Al-Dossary, R. I. R. Hamzah, and D. Mba, "Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing," *Appl. Acoust.*, vol. 70, no. 1, pp. 58–81, Jan. 2009, doi: 10.1016/j.apacoust.2008.01.005.

[5] T. Guo and Z. Deng, "An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing," *Appl. Acoust.*, vol. 127, pp. 46–62, Dec. 2017, doi: 10.1016/j.apacoust.2017.05.018.

[6] J. M. Lu, F. L. Meng, H. Shen, L. B. Ding, and S. N. Bao, "Fault diagnoise of rolling bearing based on eemd and instantaneous energy density spectrum," *Appl. Mech. Mater.*, vols. 97–98, pp. 741–744, 2011. Accessed: Sep. 18, 2020. [Online]. Available: https://www.scientific.net/AMM.97-98.741

[7] A. Kumar and R. Kumar, "Enhancing weak defect features using undecimated and adaptive wavelet transform for estimation of roller defect size in a bearing," *Tribol. Trans.*, vol. 60, no. 5, pp. 794–806, Sep. 2017, doi: 10.1080/10402004.2016.1213343.

[8] M. Singh and R. Kumar, "Thrust bearing groove race defect measurement by wavelet decomposition of pre-processed vibration signal," *Measurement*, vol. 46, no. 9, pp. 3508–3515, Nov. 2013, doi: 10.1016/j.measurement.2013.06.044.

[9] J. Saari, D. Strömbergsson, J. Lundberg, and A. Thomson, "Detection and identification of windmill bearing faults using a one-class support vector machine (SVM)," *Measurement*, vol. 137, pp. 287–301, Apr. 2019, doi: 10.1016/j.measurement.2019.01.020.

[10] A. Kumar and R. Kumar, "Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump," *Measurement*, vol. 108, pp. 119–133, Oct. 2017, doi: 10.1016/j.measurement.2017.04.041.

[11] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. M. Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renew. Energy*, vol. 90, pp. 501–512, May 2016, doi: 10.1016/j.renene.2016.01.036.

[12] Y. Zhang, X. Li, L. Gao, W. Chen, and P. Li, "Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method," *Measurement*, vol. 151, Feb. 2020, Art. no. 107232, doi: 10.1016/j.measurement.2019.107232.

[13] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images," *Appl. Acoust.*, vol. 167, Oct. 2020, Art. no. 107399, doi: 10.1016/j.apacoust.2020.107399.

[14] D. Peng, H. Wang, Z. Liu, W. Zhang, M. J. Zuo, and J. Chen, "Multi-branch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4949–4960, Jul. 2020.

[15] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016, doi: 10.1016/j.measurement.2016.07.054.

[16] M. Azamfar, J. Singh, I. Bravo-Imaz, and J. Lee, "Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis," *Mech. Syst. Signal Process.*, vol. 144, Oct. 2020, Art. no. 106861, doi: 10.1016/j.ymssp.2020.106861.

[17] Z. Chen, C. Li, and R.-V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, Oct. 2015, Art. no. 390134, doi: 10.1155/2015/390134.

[18] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[19] Z. Wu, H. Jiang, K. Zhao, and X. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, Feb. 2020, Art. no. 107227, doi: 10.1016/j.measurement.2019.107227.

[20] Z. Zhang, H. Chen, S. Li, and Z. An, "Unsupervised domain adaptation via enhanced transfer joint matching for bearing fault diagnosis," *Measurement*, vol. 165, Dec. 2020, Art. no. 108071, doi: 10.1016/j.measurement.2020.108071.

[21] Q. Li, B. Tang, L. Deng, Y. Wu, and Y. Wang, "Deep balanced domain adaptation neural networks for fault diagnosis of planetary gearboxes with limited labeled data," *Measurement*, vol. 156, May 2020, Art. no. 107570, doi: 10.1016/j.measurement.2020.107570.

[22] J. Wang, Z. Mo, H. Zhang, and Q. Miao, "Ensemble diagnosis method based on transfer learning and incremental learning towards mechanical big data," *Measurement*, vol. 155, Apr. 2020, Art. no. 107517, doi: 10.1016/j.measurement.2020.107517.

[23] J. Guo *et al.*, "Generative transfer learning for intelligent fault diagnosis of the wind turbine gearbox," *Sensors*, vol. 20, no. 5, p. 1361, Mar. 2020, doi: 10.3390/s20051361.

[24] J. Li, X. Li, D. He, and Y. Qu, "A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network," *Proc. Inst. Mech. Eng., O, J. Risk Rel.*, vol. 234, no. 1, pp. 168–182, Feb. 2020, doi: 10.1177/1748006X19867776.

[25] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019, doi: 10.1109/TSMC.2017.2754287.

[26] S. Tang, S. Yuan, and Y. Zhu, "Deep learning-based intelligent fault diagnosis methods toward rotating machinery," *IEEE Access*, vol. 8, pp. 9335–9346, 2020, doi: 10.1109/ACCESS.2019.2963092.

[27] C. Chen, Z. Li, J. Yang, and B. Liang, "A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 5622–5626, doi: 10.1109/CCDC.2017.7978168.

[28] G. Xu, M. Liu, Z. Jiang, W. Shen, and C. Huang, "Online fault diagnosis method based on transfer convolutional neural networks," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 509–520, Feb. 2020, doi: 10.1109/TIM.2019.2902003.

[29] A. Duan, L. Guo, H. Gao, X. Wu, and X. Dong, "Deep focus parallel convolutional neural network for imbalanced classification of machinery fault diagnostics," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8680–8689, Nov. 2020, doi: 10.1109/TIM.2020.2998233.

[30] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658–2669, Jun. 2020, doi: 10.1109/TIM.2019.2925247.

[31] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3325–3333, Jun. 2020, doi: 10.1109/TIM.2019.2933119.

[32] M. Sohaib and J.-M. Kim, "Fault diagnosis of rotary machine bearings under inconsistent working conditions," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3334–3347, Jun. 2020, doi: 10.1109/TIM.2019.2933342.

[33] P. P. Banik, R. Saha, and K.-D. Kim, "An automatic nucleus segmentation and CNN model based classification method of white blood cell," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113211, doi: 10.1016/j.eswa.2020.113211.

[34] R. Kumar and M. Singh, "Outer race defect width measurement in taper roller bearing using discrete wavelet transform of vibration signal," *Measurement*, vol. 46, no. 1, pp. 537–545, Jan. 2013, doi: 10.1016/j.measurement.2012.08.012.
[35] *Download a Data File | Bearing Data Center*. Accessed: Jul. 2, 2020. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/download-data-file
[36] B. Yang, Y. Lei, F. Jia, and S. Xing, "An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings," *Mech. Syst. Signal Process.*, vol. 122, pp. 692–706, May 2019, doi: 10.1016/j.ymssp.2018.12.051.
[37] F. Li, T. Tang, B. Tang, and Q. He, "Deep convolution domain-adversarial transfer learning for fault diagnosis of rolling bearings," *Measurement*, vol. 169, Feb. 2021, Art. no. 108339, doi: 10.1016/j.measurement.2020.108339.
[38] X.-G. Shang and W.-S. Jiang, "A note on fuzzy information measures," *Pattern Recognit. Lett.*, vol. 18, no. 5, pp. 425–432, May 1997, doi: 10.1016/S0167-8655(97)00028-7.

Anil Kumar received the Ph.D. degree in mechanical engineering from the Sant Longowal Institute of Engineering and Technology, Longowal, India, in 2017.

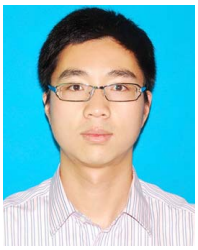
He has been with Amity University Uttar Pradesh, Noida, India, since 2016. He has authored over 20 research papers in Science Citation Index (SCI) journals. His current research interests include fault diagnosis of mechanical components, vibration and acoustic signal processing, and artificial intelligence.

Govind Vashishtha received the B.Tech. degree in mechanical engineering from Uttar Pradesh Technical University, Lucknow, India, in 2013, and the M.Tech. degree in manufacturing system engineering from the Sant Longowal Institute of Engineering and Technology, Longowal, India, in 2016, where he is currently pursuing the Ph.D. degree.

His current research interests include signal processing, condition monitoring, fault diagnosis, vibration analysis, acoustic signal processing, machine learning, and nonconventional machining processes.

C. P. Gandhi received the Ph.D. degree in mathematics from Guru Nanak Dev University, Amritsar, India, in 2011.

He is currently working as a Professor cum Head with the Department of Mathematics, Rayat Bahra University, Mohali, Chandigarh. His current research interests include information theory, neutrosophic entropy, bearing and rotor defects, and fault diagnosis.



Yuqing Zhou received the B.S. degree in computing science from Jiangnan University, Wuxi, China, in 2005, the M.S. degree from Southeast University, Nanjing, China, in 2008, and the Ph.D. degree in mechanical engineering from the Zhejiang University of Technology, Hangzhou, China, in 2019.

He is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China. He has authored more than 30 peer-reviewed international journal papers. His current research interests include tool condition monitoring, mechanical fault diagnosis, machine learning, and quality engineering.

Adam Glowacz received the Ph.D. degree in computer science from the AGH University of Science and Technology, Krakow, Poland, in 2013.

He works at the AGH University of Science and Technology as an Associate Professor, since 2020. He has authored/coauthored over 143 scientific papers (86 papers are indexed by Web of Science). He has H-index = 24 and 1445 citations from Web of Science (1235 without self-citations) and also authored over 400 reviews of scientific papers at the international conferences. His current research interests include signal processing, image processing, acoustic/vibration analysis, fault diagnosis, pattern recognition, and artificial intelligence.

Jiawei Xiang (Member, IEEE) received the Ph.D. degree in mechanical engineering from the School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2006.

He is currently the Dean and Full-Time Professor with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China. He has authored more than 150 peer-reviewed international journal papers. His current research interests include finite element method/boundary element method, fault diagnosis, monitoring and diagnosis software, and instrument development.