IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 31, 2021, accepted June 10, 2021, date of publication June 16, 2021, date of current version June 25, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3089698*

# Bearing Fault Detection Using Scalogram and Switchable Normalization-Based CNN (SN-CNN)

**DHIRAJ NEUPANE, YUNSU KIM, AND JONGWON SEOK**
Department of Information and Communication Engineering, Changwon National University, Changwon 51140, South Korea
Corresponding author: Jongwon Seok (jwseok@changwon.ac.kr)

This work was supported by the Smart Manufacturing Innovation Leaders Program through the Ministry of the Trade, Industry and Energy (MOTIE), Korea.

**ABSTRACT** Bearings play a vital role in all rotating machinery, and their failure is one of the significant causes of machine breakdown leading to a profound loss of safety and property. Therefore, the failure of rolling element bearings should be detected early while the machine fault is small. This paper presents the model that detects bearing failures using the continuous wavelet transform and classifies them using a switchable normalization-based convolutional neural network (SN-CNN). State-of-the-art accuracy was achieved with the proposed model using the Case Western Reserve University (CWRU) bearing dataset, which serves as the primary dataset for validating various algorithms for bearing failure detection. Batch normalization techniques were also employed and compared to the proposed model. The spectrogram images were also used as input for further comparison. Using switchable normalization, the proposed model achieved the testing accuracy in between 99.44% and 100% for different batch sizes and datasets.

**INDEX TERMS** Bearing fault detection, deep learning, CWRU dataset, switchable normalization, scalogram, convolutional neural network.

## I. INTRODUCTION

With the rapid advancement in science and technology, the use of electric machines has increased swiftly. Electric machinery is used ubiquitously in manufacturing applications. They are used daily and almost for all applications, which makes them work under unfavorable circumstances, humidity, and excessive loads, leading to motor breakdown resulting in huge maintenance loss, depreciation in production level, severe monetary losses, and possible risk of loss of lives. The rotating machines and induction motor, which are composed of numerous elements like rotor, stator, shaft, and bearings, play a crucial role in industrial systems. Bearings, also known as rolling element bearings (REBs), are the most crucial and the core component of any machinery, and their health state, i.e., healthy or faults and cracks at various locations, directly affects the performance, stability, efficiency, and lifespan of the machines [1], [2]. Bearings mainly consist of four elements: ball, inner-race (IR), outer-race (OR), and cage. Fig.1 shows the bearing elements and the Case Western Reserve University [3] bearing test rig.

Many studies [4], [5] on the possibility of failure of induction motors show that bearing damage accounts for about 30% of all failures. It is the most common cause of machine failure, eventually leading to serious safety and property damage, machine crashes, and probable loss of life [6]. Traditionally, maintenance of these machines has always been done after a machine breakdown has occurred. Such a posterior maintenance approach generally leads to an unexpected machine breakdown resulting in financial losses and casualties. Thus, it is very important to monitor the health of the bearings according to the machine's functional status [7]. Therefore, the detection and diagnosis of rolling bearing failures has become an essential part of scientific and technological development and engineering exploration.

The condition monitoring and fault detection mechanism of rolling element bearings are expected to provide information about the machine's actual operating status at any time without stopping the production line. Furthermore, to understand the processes associated with bearing faults properly, mechanical vibration signals are considered one of the most useful and productive sources of information, as they can detect, identify and differentiate between different types of faults [8], [9]. Although bearing vibration signals usually contain sufficient fault information, they are often nonlinear and nonstationary. Therefore, the extraction of signal features is an important step [10].

The associate editor coordinating the review of this manuscript and approving it for publication was Shunfeng Cheng.

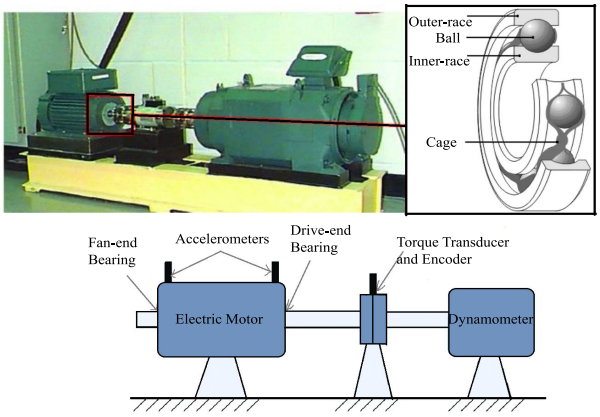


**FIGURE 1.** Bearing elements and an experimental setup of the CWRU bearing test rig for ball bearing system.

Numerous signal processing approaches in time, frequency, and time-frequency domain like measuring the peak value, peak-to-peak value, root-mean-square (RMS), and crest factor [11], skewness, kurtosis [12], and spectral kurtosis [13], impulse factor, shape factor, and clearance factor [14], wavelet transform (WT) [15], fast Fourier transform (FFT) [16], empirical mode decomposition (EMD) [17], empirical wavelet transform (EWT), ensemble empirical mode decomposition (EEMD) [18], sparse decomposition, variational mode decomposition (VMD), wavelet packet transform (WPT) [19], stochastic resonance, short-time Fourier transform (STFT) [20], Wigner-Ville distribution [21], etc., have been proposed and implemented for the bearing fault detection. However, the maximum of these methods requires different features for different types of bearing vibration data. These conventional approaches totally rely on the precise values at the fault characteristics frequencies to detect the bearing fault existence, making it crucial and troublesome to select particular features for characterizing the exact signals used for the classification.

Furthermore, machine learning (ML) algorithms are used extensively for bearing fault detection tasks. ML algorithms like artificial neural networks (ANN) [22], support vector machines (SVM) [23], principal component analysis (PCA) [24], singular value decomposition (SVD) [25], k-Nearest Neighbors (k-NN) [26] are used and have shown satisfactory performance in machinery fault detection. Nevertheless, ML algorithms cannot handle a vast amount of data. Again, the problem-solving approach of ML techniques is not end-to-end; rather, they first break down the problem statement into different parts and then combine the result [27]. The ML algorithm-based fault diagnosis methods extract fault features almost manually. However, the vibration signals of industrial equipment are often nonstationary, and the manually extracted fault features are highly dependent on the experience and prior knowledge of experts, resulting in difficulties and errors in feature extraction. Furthermore, machine learning models tend to learn only one or two data representation layers and cannot learn abundant fault information, limiting the final diagnostic accuracy. Therefore, the ML models have poor performance and cannot meet modern fault diagnosis requirements in speed and accuracy [28].

In recent years, deep learning (DL) has become an important research direction and is gradually being applied in various fields. With the advent and development of DL algorithms, the problems mentioned above have been solved significantly. Deep learning is a subfield of machine learning that can automatically learn raw data features without experts' experience and defines both higher-level and lower-level categories with higher accuracy [29]. A deep learning model consists of a multilayer neural network that extracts and learns features from deep layers of input signals. These algorithms perform outstandingly with a massive amount of data [10]. The main advantage of deep learning over traditional diagnostic algorithms is that fault features are learned through a general learning procedure, rather than being developed manually or with prior knowledge of signal processing techniques [30]. Because of the numerous advantages like end-to-end problem-solving approach, tendency to handle the complex, high-dimensional and massive amount of data, higher efficiency, reliability, and universality, DL algorithms are preferred and implemented in almost every sector. In bearing fault detection and diagnosis, different DL-based supervised, unsupervised, and transfer learning-based algorithms have been applied for fault detection and diagnosis. The deep neural structures like auto-encoders and modified auto-encoders [31], [32], deep belief networks [33], [33], generative adversarial networks [34], [35], recurrent neural networks [36], reinforcement learning [37], and convolutional neural networks (CNNs) have given highly satisfying results in this field. The domain-adaptation [38], [39] and transfer learning methods [40], [41] are also used for these tasks. These algorithms have shown their excellent performance with higher accuracy and reliability.

Talking about the CNN architectures used in bearing fault detection, both 1-D and 2-D CNNs have been employed to detect and classify bearing faults. Levent Eren proposed a 1-D CNN model in [42], which provided 97.1% accuracy. An intelligent rotating machinery fault diagnosis based on DL using data augmentation technique is proposed in [43]. The authors used two data augmentation methods and five data augmentation techniques, and the best testing accuracy obtained on the CWRU dataset was 99.91%. Similarly, a CNN-based approach with multiple sensor fusion is proposed in [44]. The average accuracy with two sensors was 99.41%, whereas that with only one sensor was 98.35%. In [45], a 2-D representation of 1-D signals is used to analyze bearing fault employing 2-D CNN, and the authors also compared their model with 1-D CNN. In [46], vibration signals, which are split into segments with the same length, are used directly as input data for CNN's deep structure. The amplitude of each sample in the vibration signal is normalized into the range [-1, 1], which becomes the corresponding pixel's intensity in the corresponding image. The accuracy, with 10 and 20 filters in the first and second layers, respectively was 96.75%. In [47],

a deep distance metric learning method in which a deep CNN is used as the leading architecture. A representation clustering algorithm is proposed to decrease the distance of intra-class variations and maximize the length of inter-class differences simultaneously. A domain adaptation method is adopted to reduce the maximum mean discrepancy between training and testing data. 99.34% accuracy was stated when the sample length is 8192. Moreover, the authors proposed the use of UNET architecture for bearing failure detection in [48]. The proposed model achieved 98.91% accuracy using vibration images as input to the UNET architecture. Again, in [49], a hierarchical adaptive deep convolutional neural network (ADCNN) is applied to the CWRU dataset, which is composed of two hierarchical layers: a fault determination layer and a fault size evaluation layer. The overall accuracy obtained is 97.7%.

### A. METHODOLOGY

The methodology implemented in this research is shown in Fig. 2. First of all, the raw bearing vibration data, originally in one-dimensional shape and stored as MATLAB files, were collected from the CWRU bearing database. Then, the scalogram images were generated using the continuous wavelet transform. The generated scalogram images were fed to the proposed switchable normalization-based convolutional neural network (SN-CNN) model. Finally, the performance analysis and evaluation using the spectrogram images as input for the same model was done. Also, for further evaluation, the batch normalization technique was used instead of switchable normalization, and the results were analyzed. The area under the curve (AUC) of each receiver operating characteristics (ROC) curve for each set of data is also tabulated and compared for further performance visualization. To further summarize the performance of this proposed model, confusion matrices are also shown for each dataset.

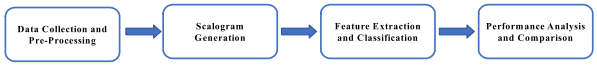


**FIGURE 2.** Methodology implemented in this research.

### B. CONTRIBUTION AND ORGANIZATION

Inspired by the widespread use of CNNs, the leading models for deep learning in computer vision, we have used the switchable normalization-based CNN model to detect and classify bearing faults using the scalogram images as input. The proposed model uses data visualization and proper feature generation characteristics of time-frequency analysis, i.e., continuous wavelet transforms (CWT), switchable normalization, which is robust, versatile, and diverse, and the proper feature extraction and classification characteristics of CNN.

The organization of this paper is as follows: It contains four sections. Section I introduces the bearing fault, conventional signal processing, and ML/DL approach used in bearing fault detection, some related works, the methodology implemented, and the contribution and organization of this research. The related theory regarding continuous wavelet transform, CNN, ReLU, dropout and SoftMax, and switchable normalization are described in section II. The following section presents the details regarding the the proposed SN-CNN model, where the data collection, pre-processing, scalogram generation, and feature extraction are described. Section IV is about the performance analysis of the proposed model. This section also shows the proposed model's performance analysis and its comparison using spectrogram images as input. Also, batch normalization instead of switchable normalization is used for further comparison in this section. The AUC of each ROC curve for each dataset is also tabulated in this section. To further summarize the performance of this proposed model, confusion matrices are also shown for each dataset. The paper is summed up in section V.

## II. RELATED THEORY

### A. CONTINUOUS WAVELET TRANSFORM

The wavelet transform is most suitable for nonstationary and nonlinear signals and can effectively increase the signal-to-noise ratio. Unlike the Fourier transform, which uses only sine and cosine functions as the basis function, the wavelet transform has an infinite set of possible basis functions. The advantage of the wavelet transform is that the window is variable. Wavelet analysis provides more accurate information about signal data than other signal analysis methods. Features localized in time and frequency can be analyzed simultaneously with high resolution [50], [51]. The two types of wavelet transform are continuous and discrete wavelet transform, and the choice of the wavelet analysis depends on what one wants to do with the data [52]. For the detailed time-frequency analysis, CWT is preferred.

Continuous and discrete wavelet transform differ in the way they discretize the scale parameter. CWT usually uses an exponential scale with a base smaller than 2, whereas the discrete wavelet transform always uses an exponential scale with a base equal to 2. Therefore, CWT discretizes the scale better than the discrete wavelet transform does. CWT is more suitable for localizing transients in nonstationary signals. Compared to the discrete wavelet method, a signal's frequency content can be detected more accurately in CWT. Again, CWT has the same time resolution as the original data in each frequency band [53]. Also, DWT is displacement dependent, while CWT is displacement invariant. CWT is the best time-frequency transform for the analysis of nonstationary signals.

The continuous wavelet transform reflects the correlation between the analyzed continuous time-domain signal $x(t)$ and a wavelet function, which is defined by the following equation [54]

$$Cw(a,b) = \int_{-\infty}^{+\infty} x(t)\,\psi_{a,b} * dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\,\psi^* \frac{(t-b)}{a} dt, \quad (1)$$

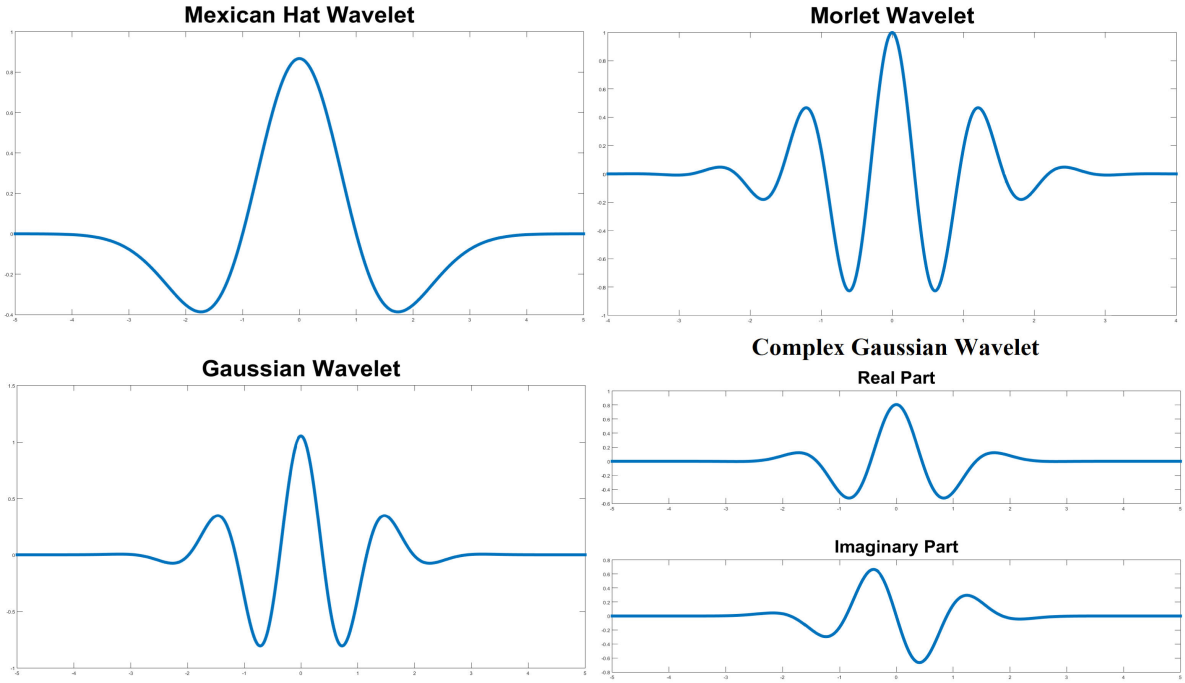


**FIGURE 3.** Different types of continuous wavelets.

where, $Cw\,(a, b)$ is the function of the parameters $a$ and $b$; $\psi$ is the continuous mother wavelet, which gets scaled by a factor of $a$ and translated by a factor of $b$. The scaling and translation factors' values are continuous, which means that there can be an infinite number of wavelets. $\psi^*\,(t)$ is the complex conjugate of the analyzing mother wavelet $\psi\,(t)$. Factor $1/\sqrt{a}$ is an energy-normalized coefficient. Besides, the wavelet function must meet the following mathematical criteria in order to be classified as a valid ground wavelet:

- The wavelet must have a finite amount of energy

$$E = \int_{-\infty}^{+\infty} |\psi\,(t)\,|^2 dt < \infty, \tag{2}$$

- $$C_\psi = \int_{-\infty}^{+\infty} \frac{|\varphi\,(w)\,|^2}{w} dw < \infty \tag{3}$$

where,

$$\varphi\,(w) = \int_{-\infty}^{+\infty} \psi(t) e^{-iwt} dt \tag{4}$$

is the Fourier transform of the $\psi(t)$ function, and $w = 2\pi f$ is the circular frequency [55].

The most used wavelets are Mexican-Hat, Morlet, Gaussian, and complex Gaussian wavelets in the continuous wavelet family. The illustrative diagrams of these wavelets are shown in Fig. 3. Mexican Hat wavelets have a very convenient analytical form for performing computations and allow for fast algorithms. They are obtained by applying the Laplacian operator to a Gaussian function and are therefore suitable for detecting Gaussian structures. Similarly, Gaussian wavelets are used in many applications. A complex (or analytic) wavelet is a function whose spectrum contains only positive frequencies. Complex wavelets respond only to the non-negative frequencies of a given signal, so they have fewer oscillatory transformations than the real wavelets. This property is a real advantage in detecting and tracking the instantaneous frequencies in a signal. The Morlet wavelet is also a wavelet with a complex exponent (carrier) multiplied by a Gaussian window (envelope). These wavelets are closely related to human perception (both auditory and visual). This wavelet family provides an intuitive bridge between frequency and time information that can clarify the interpretation of complex spectra obtained by Fourier transforms. It can also be used for bearing fault analysis.

A CWT implemented on a signal produces a spectrum of time-scale and amplitude called a scalogram. The scalogram is a plot of the absolute value of the CWT signal as a time and frequency function. Scalograms, which are obtained as a result of CWT in the form of a two-dimensional function $E(a, b) = |Cw(a, b)|^2$, present energy distribution of signals for the parameters $a$ and $b$ [55]. Since the spectrogram is obtained by windowing the input signal by shifting a window of constant length in time and frequency, Scalograms are more useful than spectrograms for analyzing real-world signals with features that occur at different scales.

In CWT, the basis is set to a multiple of 2; for example, $2^{1/v}$, where $v$ is a parameter known as the 'voices per octave.' This term is often used to refer to the number of wavelet filters per octave. "Voices are defined to be the scale levels between successive octaves, uniformly distributed in a multiplicative sense" [56]. In an audio or music signal, the number of octaves determines the range of frequencies to be analyzed, and the number of voices per octave determines the number

of samples (scale) in that range [57]. Different scales are obtained by raising this base scale to positive integer powers. The reason $v$ is referred to as the number of voices per octave is that increasing the scale by an octave (a doubling) requires $v$ intermediate scales. Common values for $v$ are 10,12,14,16, and 32. The larger the value of $v$, the finer the discretization of the scale parameter. However, this also increases the computation required because the CWT must be computed for every scale [53].

It might seem confusing to have so many options for wavelet analysis, but this is, actually, one of the strengths of wavelet analysis. Depending on the characteristics of the signal to be detected, the researcher is free to choose the wavelet that facilitates the detection of these characteristics. In this study, we have used the analytic Morlet wavelet. Various studies [58]–[60] have shown that the vibration signals of the bearing consist of periodic impulses, and the characteristics of the impact signals of the faulty bearings match well with that of the Morlet wavelet. That is why the Morlet wavelets indicate bearing defects and healthy bearing characteristics in a well-defined manner, and as a result, they are widely used in this study. Also, there is similar variability in time and frequency. The time-frequency image, a 2-D image reflecting the impact signals' energy intensity, has high instantaneous energy characteristics, making it more convenient to use the Morlet wavelet as a wavelet function. This is the reason behind choosing the Morlet wavelet in this research.

### B. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks are a special type of deep neural network (DNN) inspired by Hubel and Wiesel's work in neuroscience [61]. CNNs process the spatial correlations between neighboring pixels and represent better input images than other deep learning architectures such as autoencoders and multilayer perceptron [62]. CNNs are composed of two layers: the convolutional layer and the pooling layer. With special convolutional and pooling operations, CNNs can share parameters, and these architectures can run universally on any device. A standard CNN architecture, which is shown in Fig. 4, consists of an input (image), a feature extraction block (convolution, activation unit, pooling layer), followed by one or more fully connected layers, and finally, a classification layer [29]. The main purpose of the convolutional layer is to learn the feature representation of the input image. It is based on locally connected neurons, i.e., each neuron in the output layer receives input only from a small local group of neurons in the previous layer. It consists of several convolutional kernels that convolute with the image or the previous layer to learn different feature representations [62]. Next is the pooling layer, which acts as a down-sampling operator, combining semantically similar features into one, reducing the network's dimensionality and parameters. Two commonly used pooling operations are average pooling, which determines each patch's average value on the activation map and maximum (or maximal) pooling, determining

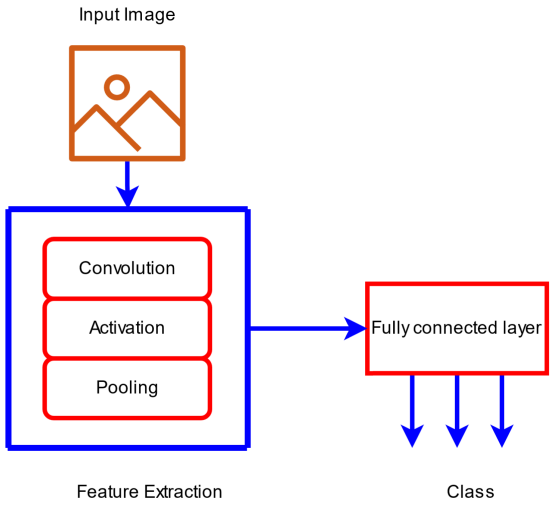


**FIGURE 4.** Common architecture of CNN.

each patch's maximum value on the feature map. In the next step, the convolutional block's output is transformed into a one-dimensional array for input to the next layer, called the fully connected layer. The last layer is the classification layer, which classifies the objects into their respective classes. CNNs have dramatically improved computer vision performance and efficiency, object recognition, natural language processing, and speech recognition. With the gradual increase in GPU production and memory, the use of CNNs in computer vision is becoming common [1], [2].

### C. ReLU, DROPOUT AND SoftMax

#### 1) ReLU

An activation function is a function that operates on the neurons of a neural network and is responsible for mapping the inputs of the neurons to their outputs [25]. The commonly used activation functions in ML/DL algorithms are sigmoid, tanh, rectified linear unit (ReLU) and its derivatives. The sigmoid and tanh functions are saturated activation functions, and the gradient vanishing problem comes up with them. On the other hand, ReLU is an unsaturated activation function, which solves gradient vanishing up to some extent and speeds up the convergence. ReLU outputs positive numbers as they are and directly sets negative numbers to zero. The calculation of the ReLU function is done after convolution. If the input is negative, ReLU is not activated at all. This function is defined as follows:

$$ReLU\ (x) = \begin{cases} x, & if\ x > 0 \\ 0, & if\ x \leq 0 \end{cases} \tag{5}$$

With the ReLU as an activation unit, a dying ReLU problem occurs, i.e., if the input goes to zero or negative, the function's slope goes to zero, and the network cannot be back-propagated and cannot be learned. This issue is solved by the Leaky ReLU activation function, which is the ReLU function derivative.

#### 2) DROPOUT

The dropout technique reduces data overfitting and generates good feature representations when training the neural network, especially with a small training dataset [42]. Overfitting of the training dataset usually results in an outstanding performance on the training data and poor performance on the test data [43]. Logically, dropout can be visualized as a set of the activated outputs of some hidden neurons to zero so that these neurons cannot participate in the forward propagation training process. Unlike in the training process, the dropout is turned off during the testing process, which indicates that all the hidden neurons participate during the testing. This process enhances the robustness and efficiency of the network [44]. Dropout makes it possible to train many networks at a reasonable time to improve the network's feature extraction capability.

#### 3) SoftMax CLASSIFIER

A SoftMax function, which is usually applied in the final layer, is a type of logistic regression used for normalizing an input value into a vector of values that follow a probability distribution whose total sums up to one [63]. This study employed the SoftMax function as the bearing health condition classifier in the proposed network. It is easy to implement and quick to compute. Let us suppose that $x^{(i)}$ be the training set with their corresponding labels $y^{(i)}$, Where $i = 1, 2, 3, \ldots, M$; $M$ is the total number of training samples. $x^{(i)} \in R^{M\times 1}$ and $y^{(i)} \in \{1, 2, 3, \ldots, \mathrm{K}\}$, where K is the number of labeled classes. For an input $x^{(i)}$, the SoftMax regression can predict the probability $P(y)^{(i)} = j|x^{(i)}$ for each label $j$, where $k = 1, 2, 3, \ldots, K$. The estimated probabilities of the input data $x^{(i)}$ belonging to each label can be obtained according to hypothesis function,

$$f_{\theta}{}^{x^{(i)}} = \begin{bmatrix} P\left(y^{(i)} = 1 \mid x^{(i)}; \theta\right) \\ P\left(y^{(i)} = 2 \mid x^{(i)}; \theta\right) \\ \vdots \\ P\left(y^{(i)} = K \mid x^{(i)}; \theta\right) \end{bmatrix} = \frac{1}{\sum_{k=1}^{K} e^{\theta_k^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_K^T x^{(i)}} \end{bmatrix}, \tag{6}$$

where, $[\theta_1, \theta_2, \ldots, \theta_K]^T$ are the parameters of the SoftMax regression model. The term

$$\theta \sum_{k=1}^{K} e^{\theta_k^T x^{(i)}}$$

normalizes the distribution, such that it sums to 1. This classifier function makes sure that the outputs are positive numbers ranging from 0 to 1, which are the probabilities for each class [64].

### D. SWITCHABLE NORMALIZATION

Normalization techniques have become an integral part of deep learning systems and play an important role in optimizing neural networks' parameters and improving generalization performance. These techniques have advanced many research areas, such as natural language processing, computer vision, and machine learning. These methods are increasingly being used to improve the stability and generalization of model learning [35]. In recent years, many normalization techniques have been developed. The standardization methods currently applied to neural networks fall into three categories.

The first is weight normalization, which normalizes the weights at the edges of the connected neurons. The second is the normalization of the activation values of layered neurons, such as batch normalization (BN), layered normalization (LN), instance normalization (IN), group normalization (GN), spectral normalization, etc. The last one is a fusion of the above methods, e.g., switchable normalization (SN) [65].

Switchable normalization is a normalization technique that can learn different normalization operations end-to-end for different normalization layers of a deep neural network. Switchable normalization combines three different statistics to estimate channel-wise, layer-wise, and mini-batch-wise with IN, LN, and BN, respectively. SN learns and switches the important weights. By design, SN can be adapted to a variety of deep neural networks and tasks. Using a one-size-fits-all normalization method may not be optimal for some tasks. For example, a combination of three normalizers is preferable for image classification and object recognition. In particular, for image classification and object recognition, SN selects BN over IN and LN. For artistic image style transfer, SN chooses IN. For neural architecture search, SN is applied to LSTM, and LN is preferred over group normalization (GN), which is a variant of IN by dividing the channel into groups [65], [66]. The intuitive graphical representation of SN is shown in Fig. 5. The feature map's size in each subplot is N × C × H × W, where N is the batch axis, C is the channel axis, and (H, W) are spatial axes. BN normalizes the NHW on the batch, IN normalizes the HW on the image pixels, LW normalizes the CHW in the channel direction; SN combines BN, LN, and IN by six weights and automatically finds a suitable normalization method during training.

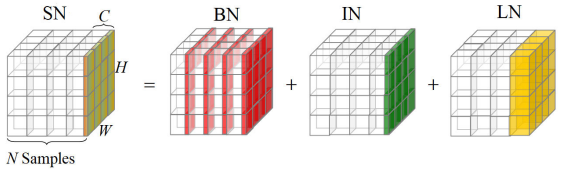


**FIGURE 5.** The intuitive graphical representation of SN.

The advantages of SN are robustness, versatility, and diversity. It can be seen from the different research that changes in mini-batch size have the most significant impact on BNs

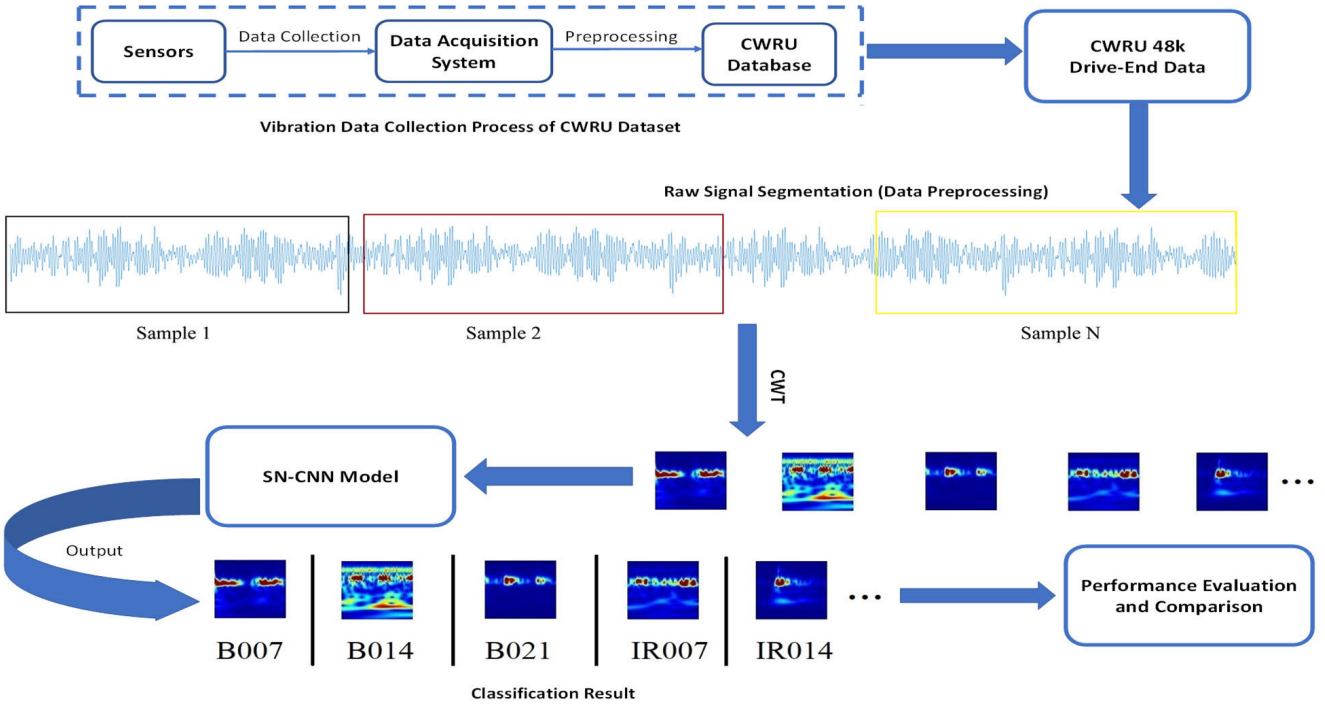


**FIGURE 6.** Flow-Chart of the proposed SN-CNN model.

because the means and variances used by BNs are counted in the mini-batch. The smaller the mini-batch, the greater the noise in estimating these statistics, and, eventually, affecting the model's performance. IN and LN have nothing to do with the mini-batch when calculating the statistics, but the lack of regularization features makes them inferior in achieving the higher accuracy. SN overcomes the problems mentioned above by learning the interaction of different normalization methods (weight coefficients). Regardless of the mini-batch size, SN can automatically learn the appropriate normalization technique and maintain a high accuracy level. In short, the smaller the minibatch, the smaller the weight coefficient of BN in SN, and the larger the weight coefficient of IN and LN; The larger the minibatch, the larger the weight coefficient of BN in SN, and the smaller the weight coefficient of IN and LN. SN is a normalization method covering the various dimensions of the feature map tensor to calculate statistical information. It does not rely on minibatch size and has good robustness to statistics in various dimensions [35], [66].

SN can be applied to many challenging benchmarks in both CNN and RNN/LSTM domains, such as image classification, object detection, segmentation, artistic image stylization, etc. SN is effective and easy to implement. In this research, we have implemented this method.

## III. SN-CNN MODEL

In this research, CNNs are used for the proper feature extraction and classification by effectively using adjacent pixel information to efficiently down-sample the image first by convolution and then using a prediction layer at the end. The CWT is used for data visualization and proper feature generation using time-frequency analysis. Similarly, the robustness, versatility, and diversity features of switchable normalization make the proposed model more effective and robust. The corresponding flow chart of the proposed method is shown in Fig. 6. The steps are listed below.

- The vibration data (48k DE data of load 1, 2, and 3 hp) were collected from the CWRU dataset and were divided into 10 classes according to the fault types. Furthermore, each class's record is further split into a certain number of samples with 2048 data points in each sample.
- The segmented raw vibration samples were then transformed into scalogram images using CWT. The scalogram image size was set to $100 \times 100 \times 3$.
- Then, the scalogram images were split to train, test and validation sets, and the corresponding label was set to each sample. They were then trained using the SN-CNN model for feature extraction and classification.
- The next step is to evaluate the performance and compare the performance using the batch normalization-based CNN (BN-CNN) model. Also, further evaluation is done using spectrogram images as an input of the SN-CNN model. The AUC of each ROC curve for each dataset is also tabulated. To further summarize the performance of this proposed model, confusion matrices are also shown for each dataset.

A detailed explanation of these steps is done in the subsequent sub-sections.

### A. DATA COLLECTION AND PRE-PROCESSING

The dataset used for this research is the CWRU bearing dataset [47], which is one of the most popular bearing datasets

and is publicly provided by Case Western Reserve University on their website. CWRU bearing dataset is mostly used for the fault analysis, classification, and detection of faulty machinery bearings; thus, it serves as the fundamental dataset to authenticate the different ML and DL-based algorithms' performance [67].

The bearing test rig arrangement used in obtaining the CWRU bearing data is shown in Fig. 1. It consists of a 2 hp Reliance electric induction motor, a dynamometer, a torque transducer, and control electronics, which is not shown in the figure. Acceleration data is collected from many sensors placed at different places. The data were collected for normal bearings, single-point drive-end (DE), and fan-end (FE) defects. This dataset consists of 161 records grouped into four classes: 48k normal-baseline, 48k drive-end fault, 12k drive-end fault, and 12k fan-end fault [68]. Electro-discharge machining was used to inject the single point faults to the test bearings with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils. One mil is equal to 0.001 inches. Vibration data was recorded for motor loads of 0 to 3 horsepower, with motor speeds of 1720 to 1797 rpm after the faulty bearings were reinstalled into the test motor [3]. For the names of the data files, the first letter represents fault position, the next three numbers signify fault diameters, and the last number denotes bearing loads. For example, the data file 'B007_0' contains the ball bearing fault data, which has the fault of diameter 0.007 inches operated under the motor load of 0 hp. Similarly, the data file 'OR014@6_1' contains the fault data of outer-race fault of diameter 0.014 inches when the load was centered (fault in 6 o'clock position) and operated under the motor load of 1 hp [1].

We have used the 48k DE fault data and 48k normal-baseline data for the experiment in this research. The fault types used are ball fault, inner-race fault, and outer-race fault for a motor load of 1 hp, 2 hp, and 3 hp for each fault type. Each fault type is further categorized into the respective fault of 7 mils, 14 mils, and 21 mils. We also have used a normal-baseline (healthy) bearing of 1 hp, 2 hp, and 3 hp load. The total no. of the dataset used is four, and each set is divided into 10 classes. Datasets are named as dataset A, dataset B, dataset C, and dataset D. Dataset A contains the 48k DE data of load 1 hp, dataset B contains 48k DE data of load 2, dataset C contains that of load 3 hp, and dataset D is the combination of all datasets A, B, and C. It might look illogical to combine the data of different loads; however, 'load' is virtually meaningless in bearing fault detection and diagnosis because of a lack of a mechanism to convert the torque to a radial load-borne by the bearings. The main consequence of the motor load is on the shaft speed, which is declined by approximately 4% in the maximum load (3 hp) case, and this would have minimal effect on the diagnosis of the datasets [6], [47]. For the data pre-processing, each record of raw vibration datasets A, B, and C is further divided into $N$ samples with 2048 data points in each sample. The value of $N$ for each class of dataset A is 186, and that for B and C is 236. Since enough data for deep learning approaches is a must, we have applied flipping as the data augmentation tool to generate new data. On flipping the samples of each class of each dataset, we doubled the number of samples in them. Thus, the total number of samples in each class of dataset A is 372, and that of datasets B and C is 472. Table 1 shows the class labels and the respective bearing fault conditions used in this research.

**TABLE 1. Class labels and the respective bearing fault conditions.**

| Class | Fault Type | Class | Fault Type |
|---|---|---|---|
| 1 | B007 | 6 | IR021 |
| 2 | B014 | 7 | Normal (Healthy) |
| 3 | B021 | 8 | OR007@6 |
| 4 | IR007 | 9 | OR014@6 |
| 5 | IR014 | 10 | OR021 |

### B. SCALOGRAM GENERATION

After the pre-processing, the next step is to generate scalogram images for each sample. For this, we used the analytical Morlet wavelet of the CWT family. The *cwtfilterbank* function of MATLAB is used to generate the scalogram images. The vibration signal length or data points in a sample is 2048. The number of voices per octave, $v$, is 12. The generated scalograms are reshaped to size $100 \times 100 \times 3$, which is the input to the SN-CNN model for feature extraction and classification. Fig. 7 shows the scalogram images of a sample of 10 different bearing health conditions.

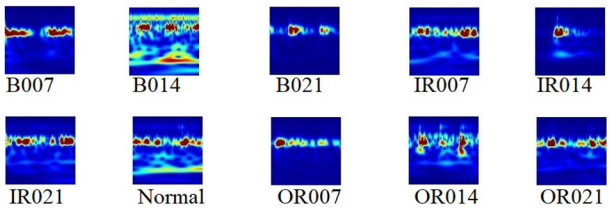


**FIGURE 7. Scalogram images of ten different health conditions.**

### C. FEATURE EXTRACTION AND CLASSIFICATION

For the feature extraction and classification, we used the SN-CNN model. The generated scalogram images of each dataset are split into a train-test ratio of 0.8:0.2. Again, 10% of the train set is further divided into a validation set. The number of classes for each dataset is 10. Table 2 shows the length of the train, test, and validation set for each dataset used in the research.

The SN-CNN model used is an 8-layer DL network. The network architecture contains five feature detection or feature extraction blocks. Each feature extraction block consists of convolutional layers, ReLU activation unit, switchable normalization, and max-pooling layers. All the 5 feature

**TABLE 2.** Length of train, test, and validation set used in the research.

| Dataset | Total Samples | Train set | Test Set | Validation Set |
|---|---|---|---|---|
| 48k DE Load1 (A) | 3720 | 2700 | 720 | 300 |
| 48k DE Load2 (B) | 4720 | 3410 | 940 | 370 |
| 48k DE Load3 (C) | 4720 | 3410 | 940 | 370 |
| D (A+B+C) | 13160 | 9510 | 2600 | 1050 |

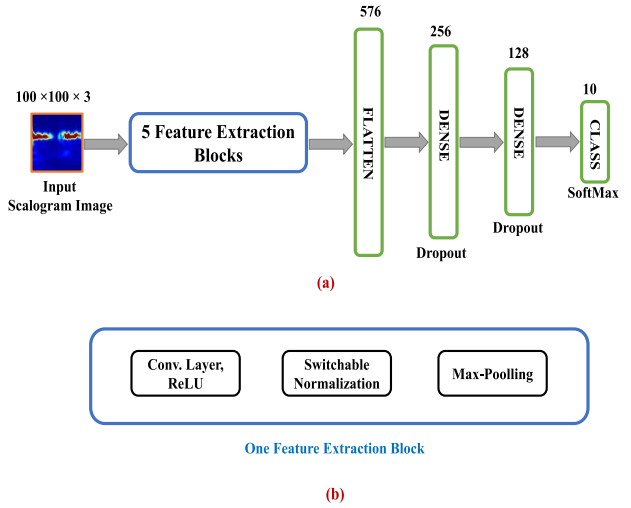


**FIGURE 8.** Network architecture of SN-CNN model.

extraction blocks contain convolutional filters of size 3 × 3, max-pooling layer of size 2 × 2, stride of size 1, ReLU activation unit, 'he_normal' kernel initialization; however, the number of layers or the length of filters differ between them. The no. of layers used in five blocks is 64, 32, 64, 32, 64, respectively. After that, two dense layers of sizes 256 and 128 were used. We have applied the dropout technique in this research to avoid the complex co-adaptations on the training data and prevent the extractions of the same features redundantly. The dropout of 0.2 was implemented in each dense layer. Finally, the classification layer for labeling ten classes was used with SoftMax as an activation function. The learning rate was set to 0.001, Adam was used as the optimizer. The network architecture of the proposed SN-CNN model is shown in Fig. 8. Moreover, information regarding the number of parameters in each layer is shown in Table 3. The total number of parameters used in the network is 258664.

## IV. PERFORMANCE ANALYSIS

The result obtained after training the SN-CNN model is quite promising. The model showed state-of-the-art accuracy performance. The training, testing, and validation accuracy for a batch size of 32 are shown in Table 4, and the graph representing the training, testing, and validation accuracy is shown in Fig 9. Also, further evaluation is done using spectrogram images as an input of the SN-CNN model. Again, the model's performance is compared using the batch-norm technique instead of the switch-norm technique. The AUC of each ROC curve for each dataset is also tabulated. The following section describes the comparison and performance analysis of the proposed model.

**TABLE 3.** Network layers and parameters of SN-CNN model.

| Layers | SN-CNN Model | |
|---|---|---|
| | Output shape | Parameters |
| Input | (None, 100, 100, 3) | 0 |
| Conv2D | (None, 100, 100, 64) | 1792 |
| SN | (None, 100, 100, 64) | 262 |
| MaxPool2D | (None, 50, 50, 64) | 0 |
| Conv2D | (None, 50, 50, 32) | 18464 |
| SN | (None, 50, 50, 32) | 134 |
| MaxPool2D | (None, 25, 25, 32) | 0 |
| Conv2D | (None, 25, 25, 64) | 18496 |
| SN | (None, 25, 25, 64) | 262 |
| MaxPool2D | (None, 12, 12, 64) | 0 |
| Conv2D | (None, 12, 12, 32) | 18464 |
| SN | (None, 12, 12, 32) | 134 |
| MaxPool2D | (None, 6, 6, 32) | 0 |
| Conv2D | (None, 6, 6, 64) | 18496 |
| SN | (None, 6, 6, 64) | 262 |
| MaxPool2D | (None, 3, 3, 64) | 0 |
| Flatten | (None, 576) | 0 |
| Dense Layer 1 | (None, 256) | 147712 |
| Dense Layer 2 | (None, 128) | 32896 |
| Classification Layer | (None, 10) | 1290 |
| | **Total Parameters: 258664** | |

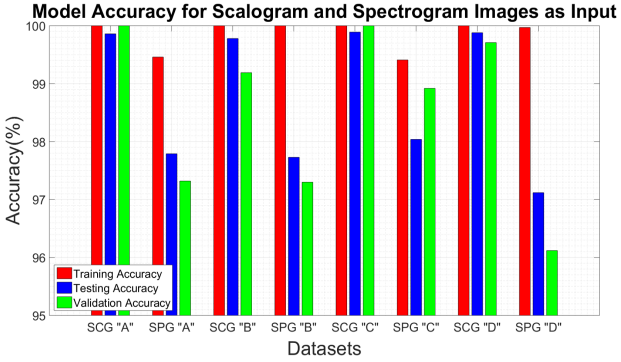


**FIGURE 9.** Performance evaluation for scalogram and spectrogram Images as input.

**TABLE 4.** SN-CNN model accuracy for scalogram and spectrogram images as input.

| Input | Dataset | Training Accuracy | Testing Accuracy | Validation Accuracy |
|---|---|---|---|---|
| Scalogram | A | 100% | 99.86% | 100% |
| Spectrogram | A | 99.46% | 97.79% | 97.32% |
| Scalogram | B | 100% | 99.78% | 99.19% |
| Spectrogram | B | 100% | 97.73% | 97.30% |
| Scalogram | C | 100% | 99.89% | 100% |
| Spectrogram | C | 99.41% | 98.04% | 98.92% |
| Scalogram | D | 100% | 99.88% | 99.71% |
| Spectrogram | D | 99.97% | 97.12% | 96.12% |

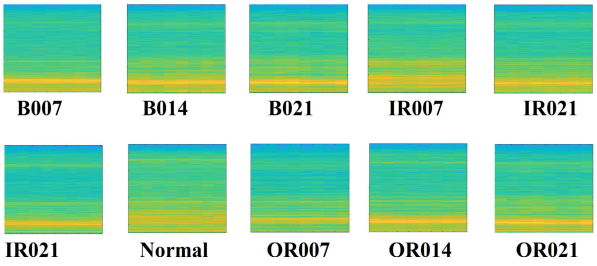


**FIGURE 10.** Spectrogram images of ten different health conditions.

### A. COMPARISON USING SPECTROGRAM IMAGES AS INPUT

To compare the performance of the proposed model, we used the spectrogram images as input and trained them in the same model. All the processes were the same except for the input data. We generated the spectrogram images of each sample of data points 2048. Fig. 10 shows the spectrogram images of each sample of 10 different bearing health conditions. The data augmentation technique used is also the same. All the other procedures and the model's parameters are the same. The result obtained using spectrogram images as input is shown in Table 4. The graph of training, testing, and validation accuracy is shown in Fig. 9, where A, B, C, and D are the datasets used. 'SCG' and 'SPG' are the abbreviations for scalogram and spectrogram images used as input.

### B. COMPARISON USING BATCH NORMALIZATION FOR SCALOGRAM IMAGES

We have used the batch normalization technique instead of switchable normalization in the same model for further comparison, and analyzed the results obtained. For these tasks, different tests were done with different batch sizes. We used the batch sizes of 16, 32, and 64 for each dataset and evaluated the performance. The result is tabulated in Table 5. Here, the SN-CNN and Batch Normalization-based CNN (BN-CNN) model's input is scalogram images. Fig. 11, Fig. 12, and Fig. 13 show the SN-CNN and BN-CNN model's performance evaluation for the batch size of 16, 32, and 64, respectively. A, B, C, and D represent the datasets used.

**TABLE 5.** The efficiency of SN-CNN and BN-CNN model using scalogram as input.[1]

| DS/BS | Accuracy (SN-CNN) | | | Accuracy (BN-CNN) | | |
|---|---|---|---|---|---|---|
| | Train | Test | Valid. | Train | Test | Valid. |
| A/16 | 100% | 99.58% | 100% | 99.96% | 98.21% | 94.33% |
| B/16 | 100% | 99.78% | 100% | 99.65% | 97.60% | 98.68% |
| C/16 | 100% | 100% | 100% | 99.68% | 97.90% | 98.42 |
| D/16 | 100% | 99.92% | 100% | 99.83% | 98.18% | 98.11% |
| A/32 | 100% | 99.72% | 100% | 99.96% | 98.22% | 94.67% |
| B/32 | 99.73% | 99.79% | 99.73% | 99.68% | 97.91% | 98.38% |
| C/32 | 100% | 99.79% | 100% | 99.39% | 97.81% | 98.42% |
| D/32 | 100% | 99.73% | 99.52% | 99.74% | 97.89% | 97.26% |
| A/64 | 100% | 99.44% | 99.33% | 99.93% | 98.63% | 94.67% |
| B/64 | 100% | 99.57% | 100% | 99.65% | 98.01% | 98.68% |
| C/64 | 100% | 99.68% | 99.46% | 99.65% | 98.22% | 98.42% |
| D/64 | 100% | 99.84% | 99.52% | 99.84% | 98.14% | 97.29% |

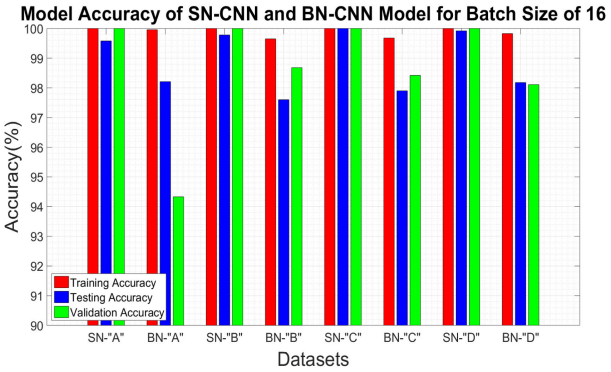


**FIGURE 11.** Performance evaluation of SN-CNN and BN-CNN model for batch size of 16.

### C. ROC CURVE, AUC, AND CONFUSION MATRIX

The receiver operating characteristics curve or ROC graphs are a method for visualizing, organizing, and selecting classifiers based on their performance. These curves are graphical plots that illustrate a classification system's diagnostic ability. ROC graphs have long been used in signal detection theory to represent the tradeoff between a classifier's hit rate and false alarm rate. ROC is the probability curve, and AUC is the degree or measure of separability. It indicates the degree to which the model can distinguish between classes. The Higher

[1] DS/BS: Dataset/ Batch Size Valid.: Validation

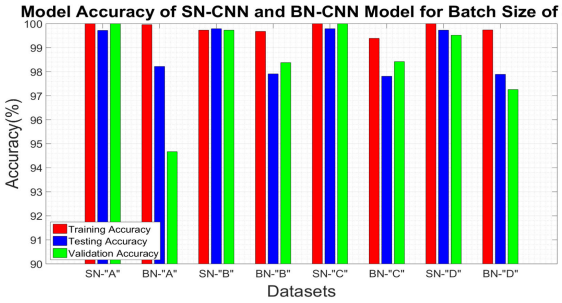


**FIGURE 12.** Performance evaluation of SN-CNN and BN-CNN model for batch size of 32.

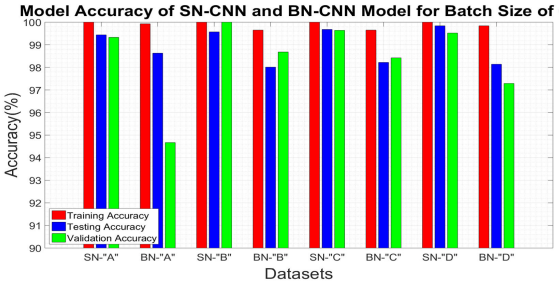


**FIGURE 13.** Performance evaluation of SN-CNN and BN-CNN model for batch size of 64.

the AUC, the better the model distinguishes faulty bearing classes [69], [70].

The area under the ROC curve of each class of each dataset is tabulated in Table 6. From the table, it is clear that most of the area under the ROC curve is found to be 1 for the SN-CNN models. Whereas, for the BN-CNN models, none of the AUC is 1. The batch size in determining the ROC curve is set to default to 32. The ROC curves are not shown in this paper as there are 80 images for four datasets, using BN-CNN and SN-CNN moel. However, if anyone is interested in visualizing them, we have uploaded them in our GitHub repository (link: https://github.com/dhirajneupane/BFD-SN-CNN).

The confusion matrix is a summary of the classification performance of a classifier on test data. It is a two-dimensional matrix indexed by the classifier's true class in one dimension and the class assigned by the classifier in the other dimension [71]. In the confusion matrix, each column contains a value representing an instance of the predicted class, and each row contains a value representing an instance of the actual class. The name comes from the fact that the system can easily check whether the system is confused among the classes or not [72].

The confusion matrix obtained for the proposed model has ten rows and ten columns. Fig. 14, Fig. 15, Fig. 16, and Fig 17 represent the comparison of the confusion matrix obtained by the BN-CNN and SN-CNN model for the datasets 'A', 'B', 'C' and 'D' respectively.

### D. RESULT ANALYSIS

The results show that the proposed model achieves state-of-art accuracy. With the use of scalogram images as input, the information regarding the health state of the bearings is more defining, features are more distinctive than using the spectrogram images as input. Moreover, with switchable normalization, the proposed system achieved the testing

**TABLE 6.** AUC of each class obtained using SN-CNN and BN-CNN model.

| Dataset | SN-CNN Model | | BN-CNN Model | |
|---|---|---|---|---|
| | Class | AUC | Class | AUC |
| A | B007 | 1.00 | B007 | 0.9822 |
| | B014 | 1.00 | B014 | 0.9881 |
| | B021 | 1.00 | B021 | 0.9875 |
| | IR007 | 1.00 | IR007 | 0.9880 |
| | IR014 | 1.00 | IR014 | 0.9876 |
| | IR021 | 1.00 | IR021 | 0.9883 |
| | Normal | 1.00 | Normal | 0.9895 |
| | OR007 | 1.00 | OR007 | 0.9876 |
| | OR014 | 1.00 | OR014 | 0.9873 |
| | OR021 | 1.00 | OR021 | 0.9866 |
| B | B007 | 1.00 | B007 | 0.9942 |
| | B014 | 1.00 | B014 | 0.9926 |
| | B021 | 1.00 | B021 | 0.9954 |
| | IR007 | 1.00 | IR007 | 0.9952 |
| | IR014 | 1.00 | IR014 | 0.9958 |
| | IR021 | 1.00 | IR021 | 0.9948 |
| | Normal | 1.00 | Normal | 0.9920 |
| | OR007 | 1.00 | OR007 | 0.9973 |
| | OR014 | 1.00 | OR014 | 0.9963 |
| | OR021 | 1.00 | OR021 | 0.9913 |
| C | B007 | 1.00 | B007 | 0.9872 |
| | B014 | 1.00 | B014 | 0.9935 |
| | B021 | 1.00 | B021 | 0.9943 |
| | IR007 | 1.00 | IR007 | 0.9930 |
| | IR014 | 1.00 | IR014 | 0.9930 |
| | IR021 | 1.00 | IR021 | 0.9930 |
| | Normal | 1.00 | Normal | 0.9979 |
| | OR007 | 1.00 | OR007 | 0.9931 |
| | OR014 | 1.00 | OR014 | 0.9946 |
| | OR021 | 1.00 | OR021 | 0.9926 |
| D | B007 | 0.9999 | B007 | 0.9909 |
| | B014 | 1.00 | B014 | 0.9932 |
| | B021 | 1.00 | B021 | 0.9931 |
| | IR007 | 1.00 | IR007 | 0.9919 |
| | IR014 | 1.00 | IR014 | 0.9926 |
| | IR021 | 1.00 | IR021 | 0.9910 |
| | Normal | 1.00 | Normal | 0.9914 |
| | OR007 | 1.00 | OR007 | 0.9911 |
| | OR014 | 0.9999 | OR014 | 0.9941 |
| | OR021 | 1.00 | OR021 | 0.9920 |

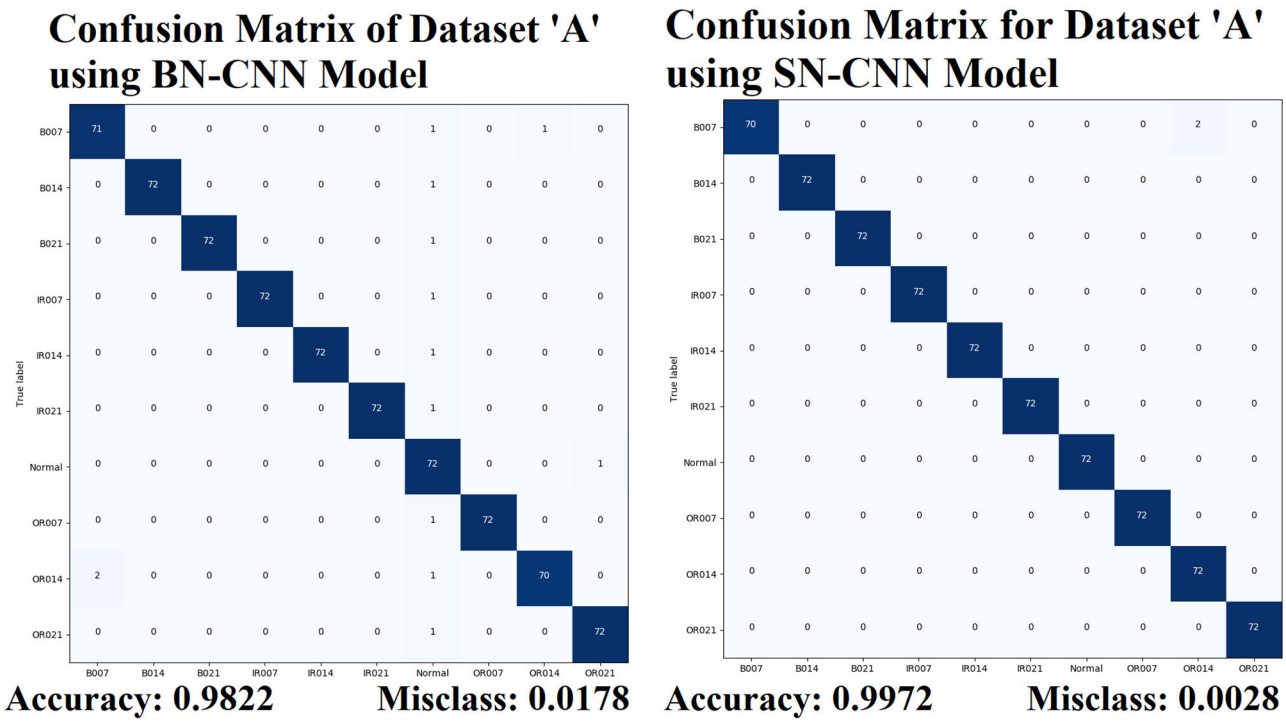


**FIGURE 14.** Confusion matrix of dataset 'A' using BN-CNN model (left) and SN-CNN model (right).

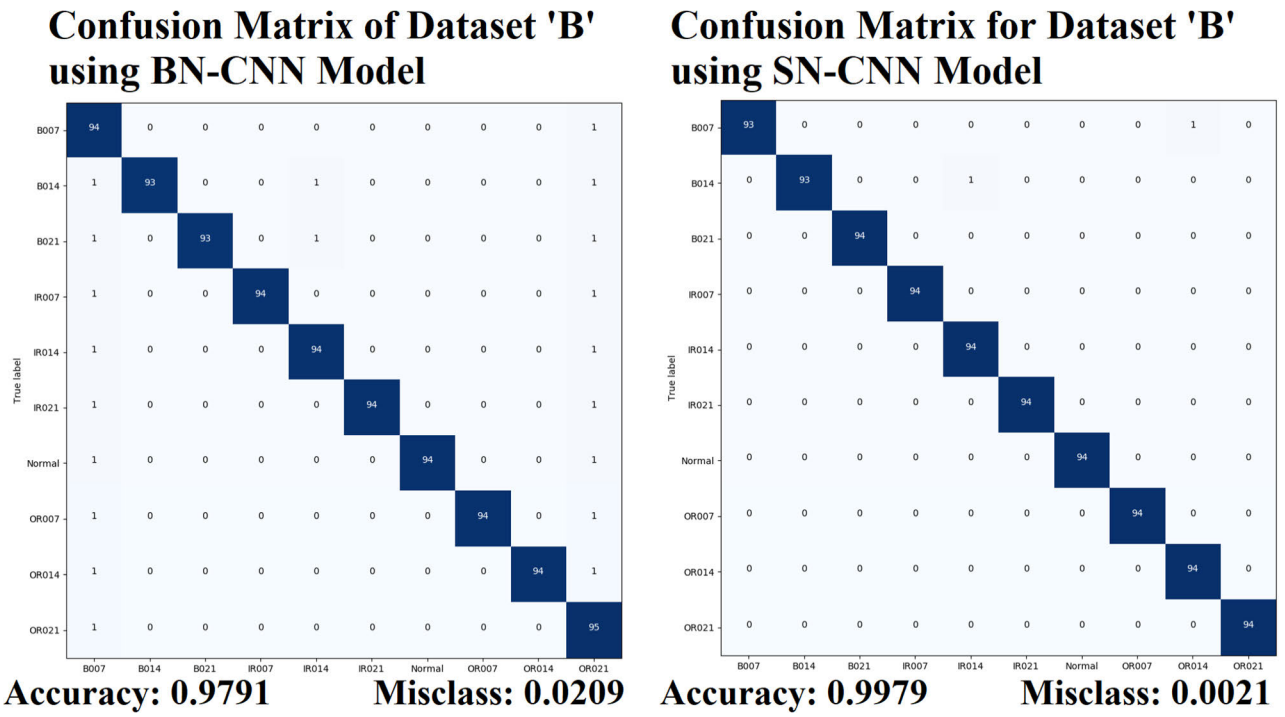


**FIGURE 15.** Confusion matrix of dataset 'B' using BN-CNN model (left) and SN-CNN model (right).

accuracy in between 99.44% and 100% for different batch sizes and datasets. Furthermore, the confusion matrices of BN-CNN and SN-CNN models also clarify that switchable normalization is more robust and reliable. In the training and testing of a large-scale dataset, the difference between switch-norm and batch-norm is visualized more.

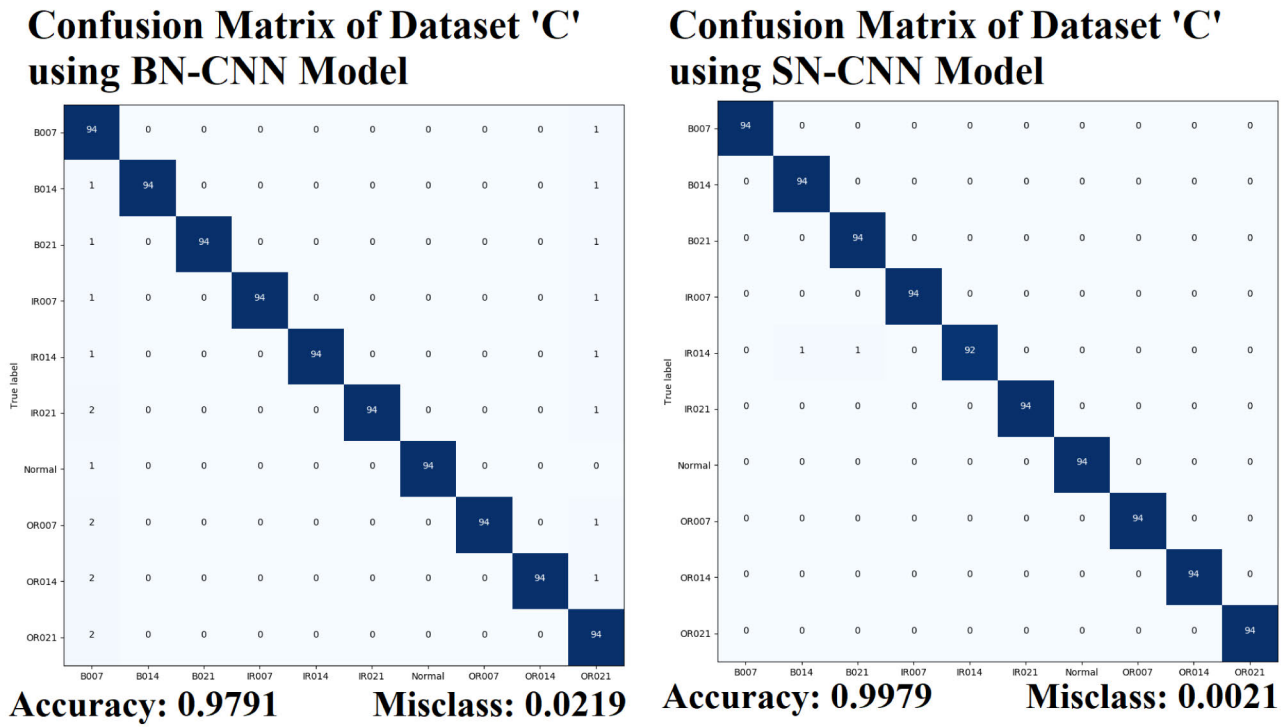


**FIGURE 16.** Confusion matrix of dataset 'C' using BN-CNN model (left) and SN-CNN model (right).

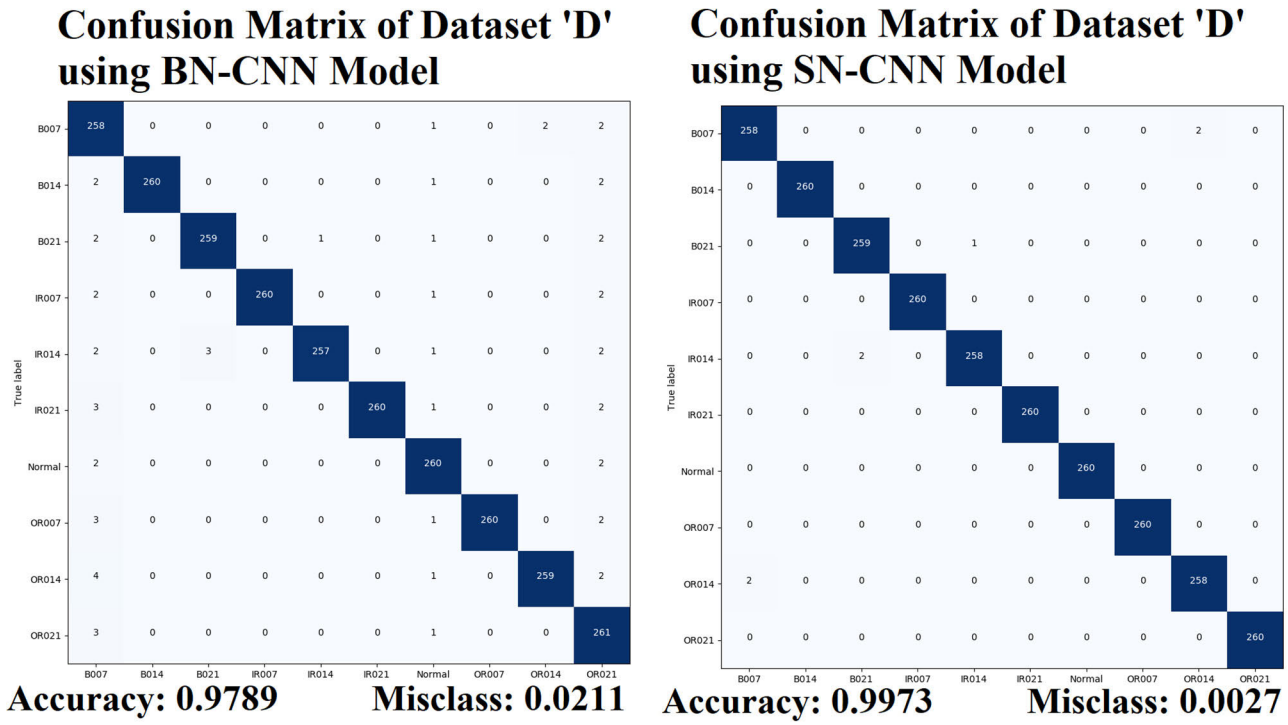


**FIGURE 17.** Confusion matrix of dataset 'D' using BN-CNN model (left) and SN-CNN model (right).

## V. DISCUSSION AND CONCLUSION

In industry 4.0, deep learning algorithms for various research applications are attracting much attention. Recently, DL models have been widely used in machine fault detection and diagnostic systems. As rapid advances in computer technology continue to be made, DL models are likely

to be a robust and attractive tool in machine fault detection and diagnostic systems. In this research, we have used the advantages of a signal processing method, i.e., continuous wavelet transforms, for feature generation from raw time-domain CWRU bearing database, a widely used dataset to detect and diagnose machine bearing failures and accepted as a standard reference for model validation, and implemented switchable normalization-based convolutional neural network (SN-CNN) for the feature extraction and classification of the normal (healthy) and faulty bearing signals. We carried our experiment on four sets of 48k drive-end bearing data. We also compared this model's performance by using the spectrogram images as input. We used the batch normalization technique in our model instead of switchable normalization and analyzed the result for further comparison. Using switchable normalization, the proposed system achieved the testing accuracy in between 99.44% and 100% for different batch sizes and datasets. The results show that the proposed system attains state-of-the-art accuracy when the scalogram images are input.

## ACKNOWLEDGMENT

The authors would like to express their sincere appreciation to Case Western Reserve University for the open-source CWRU bearing data.

## REFERENCES

[1] D. Neupane and J. Seok, "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review," *IEEE Access*, vol. 8, pp. 93155–93178, 2020.

[2] D. Neupane and J. Seok, "Deep learning-based bearing fault detection using 2-D illustration of time sequence," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2020, pp. 562–566.

[3] *Case Western Reserve University Bearing Data Center*. Accessed: Dec. 22, 2019. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/home

[4] R. N. Bell, D. W. McWilliams, P. O'Donnell, C. Singh, and S. J. Wells, "Report of large motor reliability survey of industrial and commercial installations, Part I," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 853–864, Jul. 1985.

[5] R. N. Bell, C. R. Heising, P. O'Donnell, S. J. Wells, and C. Singh, "Report of large motor reliability survey of industrial and commercial installations, Part II," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 865–872, Jul. 1985.

[6] W. Huang, J. Cheng, and Y. Yang, "Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection," *Mech. Syst. Signal Process.*, vol. 114, pp. 165–188, Jan. 2019.

[7] C. Sun, Z. Zhang, Z. He, Z. Shen, and B. Chen, "Manifold learning-based subspace distance for machinery damage assessment," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 637–649, Mar. 2016.

[8] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *J. Failure Anal. Prevention*, vol. 16, no. 2, pp. 271–284, Apr. 2016.

[9] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, no. 1, pp. 1–8, Jul. 2014.

[10] L. Yuan, D. Lian, X. Kang, Y. Chen, and K. Zhai, "Rolling bearing fault diagnosis based on convolutional neural network and support vector machine," *IEEE Access*, vol. 8, pp. 137395–137406, 2020.

[11] N. Tandon and B. C. Nakra, "Detection of defects in rolling element bearings by vibration monitoring," *J. Mech. Eng. Division*, vol. 73, pp. 271–282, Jan. 1993.

[12] D. Dyer and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *J. Mech. Des.*, vol. 100, no. 2, pp. 229–235, Apr. 1978.

[13] R. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, Apr. 1983, pp. 607–610.

[14] A. Shrivastava and S. Wadhwani, "Development of fault detection system for ball bearing of induction motor using vibration signal," *Int. J. Sci. Res.*, vol. 2, pp. 256–259, May 2013.

[15] Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *J. Sound Vib.*, vol. 253, no. 5, pp. 1087–1100, Jun. 2002.

[16] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.

[17] C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, "Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis," 2019, pp. 1–11, *arXiv:1903.06753*. [Online]. Available: http://arxiv.org/abs/1903.06753

[18] X. Wang, C. Liu, F. Bi, X. Bi, and K. Shao, "Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 581–597, Dec. 2013.

[19] C. Castejón, M. Jesús, J. J. Gómez, J. C. García-Prada, A. J. O. Nez, and H. Rubio, "Automatic selection of the WPT decomposition level for condition monitoring of rotor elements based on the sensitivity analysis of the wavelet energy," *Int. J. Acoust. Vib.*, vol. 20, no. 2, pp. 95–100, Jun. 2015.

[20] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "STFT based approach for ball bearing fault detection in a varying speed motor," in *Condition Monitoring of Machinery in Non-Stationary Operations*, T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, and M. Haddar, Eds. Berlin, Germany: Springer, 2012, doi: 10.1007/978-3-642-28768-8_5.

[21] J. H. Lee, J. Kim, and H. J. Kim, "Development of enhanced Wigner–Ville distribution function," *Mech. Syst. Signal Process.*, vol. 15, no. 2, pp. 367–398, Mar. 2001.

[22] M.-Y. Chow, P. M. Mangum, and S. O. Yee, "A neural network approach to real-time condition monitoring of induction motors," *IEEE Trans. Ind. Electron.*, vol. 38, no. 6, pp. 448–453, Dec. 1991.

[23] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, no. 1, pp. 1–9, Jan. 2014.

[24] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[25] F. Shen, C. Chen, R. Yan, and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM)*, no. 1, 2016, pp. 1–6.

[26] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4137–4145, Aug. 2013.

[27] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

[28] Z. Chen, J. Cen, and J. Xiong, "Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network," *IEEE Access*, vol. 8, pp. 150248–150261, 2020.

[29] D. Neupane and J. Seok, "A review on deep learning-based approaches for automatic sonar target recognition," *Electronics*, vol. 9, no. 11, p. 1972, Nov. 2020.

[30] Y. Yang, P. Fu, and Y. He, "Bearing fault automatic classification based on deep learning," *IEEE Access*, vol. 6, pp. 71540–71554, 2018.

[31] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.

[32] C. Li, W. Zhang, G. Peng, and S. Liu, "Bearing fault diagnosis using fully-connected winner-take-all autoencoder," *IEEE Access*, vol. 6, pp. 6103–6115, 2018.

[33] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Sep. 2015, Art. no. 115002.

[34] W. Jiang, C. Cheng, B. Zhou, G. Ma, and Y. Yuan, "A novel GAN-based fault diagnosis approach for imbalanced industrial time series," 2019, *arXiv:1904.00575*. Accessed: Jan. 7, 2020. [Online]. Available: http://arxiv.org/abs/1904.00575

[35] D. Zhao, F. Liu, and H. Meng, "Bearing fault diagnosis based on the switchable normalization SSGAN with 1-D representation of vibration signals as input," *Sensors*, vol. 19, no. 19, p. 2000, 2019.

[36] H. Jiang, X. Li, H. Shao, and K. Zhao, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Meas. Sci. Technol.*, vol. 29, no. 6, May 2018, Art. no. 065107.

[37] R. Wang, H. Jiang, X. Li, and S. Liu, "A reinforcement neural architecture search method for rolling bearing fault diagnosis," *Measurement*, vol. 154, Mar. 2019, Art. no. 107417.

[38] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020.

[39] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[40] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.

[41] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput. Appl.*, vol. 32, pp. 6111–6124, 2020, doi: 10.1007/s00521-019-04097-w.

[42] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Math. Problems Eng.*, vol. 2017, pp. 1–9, Jul. 2017.

[43] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 433–452, Feb. 2020.

[44] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.

[45] W. Zhang, G. Peng, and C. Li, "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input," in *Proc. MATEC Web Conf.*, vol. 95, 2017, pp. 1–5.

[46] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognit. Syst. Res.*, vol. 53, pp. 42–50, Jan. 2019.

[47] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, Oct. 2018.

[48] D. K. Soother, I. H. Kalwar, T. Hussain, B. S. Chowdhry, S. M. Ujjan, and T. D. Memon, "A novel method based on UNET for bearing fault diagnosis," *Comput., Mater. Continua*, vol. 69, no. 1, pp. 393–408, 2021.

[49] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Meas. J. Int. Meas. Confederation*, vol. 93, pp. 490–502, Nov. 2016.

[50] P. M. Bentley and J. T. E. McDonnell, "Wavelet transforms: An introduction," *Electron. Commun. Eng. J.*, vol. 6, no. 4, pp. 175–186, Aug. 1994.

[51] L. Saribulut, A. Teke, M. B. Latran, and M. Tümay, "Fundamentals and literature review of wavelet transform in power quality issues," *J. Electr. Electron. Eng. Res.*, vol. 5, no. 1, pp. 1–8, May 2013.

[52] MATLAB & Simulink-MathWorks. *Choose a Wavelet*. Accessed: Nov. 15, 2020. [Online]. Available: https://kr.mathworks.com/help/wavelet/gs/choose-a-wavelet.html

[53] MATLAB & Simulink-MathWorks. *Continuous and Discrete Wavelet Transforms*. Accessed: Nov. 15, 2020. [Online]. Available: https://kr.mathworks.com/help/wavelet/gs/continuous-and-discrete-wavelet-transforms.html

[54] L. Navarro, G. Courbebaisse, and M. Jourlin, "Logarithmic wavelets," in *Advances in Imaging and Electron Physics*, vol. 183. New York, NY, USA: Academic, 2014, pp. 41–98.

[55] D. Komorowski and S. Pietraszek, "The use of continuous wavelet transform based on the fast Fourier transform in the analysis of multi-channel electrogastrography recordings," *J. Med. Syst.*, vol. 40, no. 1, pp. 1–15, Jan. 2016.

[56] J. Sadowsky, "The continuous wavelet transform: A tool for signal investigation and understanding," *Johns Hopkins APL Tech. Dig.*, vol. 15, no. 4, pp. 306–318, 1994.

[57] C. Liner, "An overview of wavelet transform concepts and applications," Tech. Rep., 2010. [Online]. Available: https://www.semanticscholar.org/paper/An-overview-of-wavelet-transform-concepts-and-Liner/52372db16936b0188f5257b80ef5804fff96c411#citing-papers

[58] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 8, no. 7, p. 1102, Jul. 2018.

[59] J. Guo, X. Liu, S. Li, and Z. Wang, "Bearing intelligent fault diagnosis based on wavelet transform and convolutional neural network," *Shock Vib.*, vol. 2020, pp. 1–14, Nov. 2020.

[60] S. Guo, T. Yang, W. Gao, and C. Zhang, "A novel fault diagnosis method for rotating machinery based on a convolutional neural network," *Sensors*, vol. 18, no. 5, p. 1429, May 2018.

[61] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962.

[62] N. Ranjan, S. Bhandari, H. P. Zhao, H. Kim, and P. Khan, "City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN," *IEEE Access*, vol. 8, pp. 81606–81620, 2020.

[63] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," Dec. 2016, *arXiv:1612.02295*. [Online]. Available: http://arxiv.org/abs/1612.02295

[64] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[65] P. Luo, R. Zhang, J. Ren, Z. Peng, and J. Li, "Switchable normalization for learning-to-normalize deep representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 712–728, Feb. 2021.

[66] P. Luo, J. Ren, Z. Peng, R. Zhang, and J. Li, "Differentiable learning-to-normalize via switchable normalization," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–19.

[67] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.

[68] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.

[69] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[70] D. J. Hand, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Mach. Learn.*, vol. 45, no. 2, pp. 171–186, 2001.

[71] K. M. Ting, "Confusion matrix," in *Encyclopedia of Machine Learning and Data Mining*. New York, NY, USA: Springer, 2017, p. 260.

[72] R. M. George and J. A. Mathew, "Emotion classification using machine learning and data preprocessing approach on Tulu speech data," *Int. J. Comput. Sci. Mobile Comput.*, vol. 5, no. 6, pp. 589–600, Jun. 2016.

**DHIRAJ NEUPANE** received the bachelor's degree in electrical and electronics (communication) engineering from Kathmandu University, Nepal, in 2016. He has been working as a Graduate Research Assistant with Changwon National University, South Korea, since 2019. His research interests include machine learning, deep learning, intelligent fault diagnosis method, sonar signal processing, transfer learning, machinery fault detection and diagnosis, object detection and classification, and augmented reality.

**YUNSU KIM** received the bachelor's degree in information and communication engineering from Changwon National University, South Korea, in 2020. He is currently working as a Graduate Research Assistant with Changwon National University. His research interests include machine learning, deep learning, computer vision, face detection and recognition, audio signal processing, and bandwidth extension.

**JONGWON SEOK** received the B.S. and M.S. degrees and the Ph.D. degree in electronics from Kyungpook National University, Daegu, South Korea, in 1993, 1995, and 1999, respectively. He was a Senior Member of the Research Staff with the Electronics and Telecommunications Research Institute (ETRI) from July 1999 to February 2004, where he was engaged in the development of a multichannel audio codec system, digital broadcasting content protection and management. Since March 2004, he has been with Changwon National University, where he is currently a Professor with the Department of Information and Communication Engineering. His research interests include digital and sonar signal processing, multimedia system, and deep learning applications.