BJES Basrah Journal for Engineering Sciences

Basrah Journal for Engineering Sciences, Vol. 25, No. 1, (2025), 63-73

*Original Article*

Journal homepage: www.bjes.edu.iq
ISSN (Online): 23118385, ISSN (Print): 18146120

# Bearing Fault Diagnosis Based on Acoustic and Vibration Signals using Deep Learning Neural Network

Hussein Naser Jaber [1]*, Abdulbaseer S. Bahedh [2], Reza Ale Ali [3]

[1,2] *Department of Mechanical Engineering, College of Engineering, University of Basrah, Basrah, Iraq*
[3] *Marine Engineering Department, Khorramshahr University of Marine Science and Technology (KMSU), Khorramshahr, Iran*
*E-mail addresses: pgs.hussein.naser@uobasrah.edu.iq , abdalbaseer.baheth@uobasrah.edu.iq , reza.aleali.1001@gmail.com*

***Article Info***

***Article history:***

*Received: 20 December 2024*
*Revised: 25 February 2025*
*Accepted: 6 March 2025*
*Published: 16 August 2025*

***Keywords:***

*Acoustic emission, Vibration signal, Deep learning neural network, Kurtogram.*

https://doi.org/10.33971/bjes.25.1.9

***Abstract***

Diagnosing faults in rotary machines is critical, as early fault detection is a precise and essential task in minimizing operational risks and economic losses. Bearings are vital components in rotary machines and are subject to gradual degradation due to continuous operation. Failure to detect early damage can lead to problem escalation, resulting in severe damage and increased costs. In this study, two types of signals from rotary machines are analyzed: acoustic emission (AE) signals and vibration signals. These signals are utilized as input features for a deep learning neural network based on images, where the features are extracted using the Kurtogram, a powerful fourth-order spectral analysis tool that generalizes spectral kurtosis (SK) for a given signal. The results demonstrate that the accuracy of diagnosing the machine's operational condition whether healthy or faulty ranges from 99.2% to 100%, while the accuracy of fault classification reaches 96.6%. These findings highlight the high efficiency of this methodology in fault detection and classification, establishing it as one of the most important techniques for diagnosing faults in rotary machines.

## 1. Introduction

Vibration and (AE) signals in machines are common indicators of potential malfunctions that can adversely affect the performance of rotary equipment. Various faults can occur during mechanical operations that generate vibration and (AE) signals with distinct characteristics. Therefore, early detection and prediction of mechanical failures are vital for ensuring safety and minimizing expenses. Machine learning and signal processing technologies enable more accurate predictions of faults such as bearing defects, shaft misalignment, and disk imbalances, making proactive maintenance more feasible. Mangurul and Jung [1] used an adaptive deep convolutional neural network (ADCNN) and a 2D visualization tool to describe the health status of the bearing. Initially, wavelet packet transform (WPT) was applied to measure each sub-domain by forming the degree of defectiveness ratio (DDR), an innovative evaluation tool for diagnosing bearing faults. Qian [2] developed an adaptive overlay CNN (AOCNN) for fault diagnosis to directly process unprocessed one-dimensional vibration signals. The adaptive convolutional layer and the real-time pooling Root-Mean-Square (RMS) layer are designed to handle the transformation differences.

In contrast, the overlay layer helps alleviate the inherent marginal issues commonly encountered by Convolutional neural networks (CNNs) for diagnosing bearing faults. RUYI Huang et al. [3] using multi-stack capsules as a separation classifier to identify and separate complex faults successfully. Jiedi et al. [4] reduced the size of the computed data while preserving all the error information and create a deep neural network based on stacked sparse autoencoders using an unsupervised learning process and a fine-tuning method to detect hidden discrimination in the obtained data. Wei et al. [5] used deep (CNNs) with Wide First-layer Kernels (WDCNN). Raw vibration signals are used as input in this method, and extra input samples are generated using data augmentation. The model employs broad kernels in the first convolutional layer to extract features and decrease high-frequency noise. Sandeep et al. [6] proposed method combines the kurtogram and (CNN) for fault categorization. The kurtogram measures the time-frequency energy density dispersion, and its results serve as the CNN's input feature vector. This 2D feature vector is then fed into the CNN for fault classification. Sandeep et al. [7] used deep learning sequence models (SM) and the kurtogram. For a clear picture of the faults, the SM uses the kurtogram as a sequential data set to diagnose the problem. Wei et al. [8] used (CNN) with Training Interference (TICNN) to address the issue with fault diagnostics that may be resolved. Lu´ıs et al. [9] used a convolutional neural network on temporal input signals and combined it with additional static features predicting faulty engines. Damage prediction employs vast amounts of unbalanced data from signals associated with internal engine excitation, or structure-borne noise. It considers the imbalance of the data using informative mini batches during training. Duy and Hee [10] worked on the structure of deep CNN for diagnosing bearing faults. Vibration signals are directly employed as input data. This fault diagnosis system operates automatically and doesn't require feature extraction methods. Huaqing et al. [11] integrate multi-sensor data fusion with a bottleneck layer-optimized convolutional neural network (MB-CNN). Use a conversion approach that combines data to produce images from vibration

signals from several sensors with richer details than from a single sensor.

Zhang et al. [12] used a CNN architecture and a signal image as input to detect bearing faults intelligently. With training, CNN, a potent feature extractor and classifier, through training, CNN can discover which features are best suited for the categorization assignment. Fed with signal images, CNN is quite good at identifying the periodic features of the vibration signals. Zhuyun et al. [13] apply a novel deep learning-based approach for bearing fault diagnosis, combining CNN and Cyclic Spectral Coherence (CSCoh) to determine the fault characteristics of each fault type. (CSCoh) is first employed as a preprocessing phase, extracting several discriminative features early on. A CNN is then used for additional feature learning and categorization. Yudong Cao et al. [14] presented a Temporal Convolutional Network with Residual Self-Attention mechanism (TCN-RSA), a novel deep prognostic network for rolling bearing Remaining Useful Life (RUL) prediction under various operating circumstances. The TCN-RSA acts as a hidden feature extractor, calculating the marginal spectrum of bearing vibration signals as input. TCN-RSA learns high-level representations from the data input by stacking many temporal convolution blocks. Xiong et al. [15] have suggested a new network, WPT-CNN, for rolling bearing end-to-end intelligent defect diagnosis. WPT-CNN innovatively incorporates the Wavelet Packet Transform (WPT) for time-frequency analysis within a conventional deep neural network architecture, efficiently combining domain expertise in defect diagnostics with deep learning methods. Chen and Lee [16] discussed the analysis of vibration signals using convolutional neural networks (CNNs), incorporating tool wear detection, diagnosis of bearing faults, and evaluation of machining surface roughness. For regression and classification applications, One-dimensional CNNs (1DCNN) and two-dimensional CNNs (2DCNN) are utilized with various input types, including raw signals and time-frequency spectrum images obtained through short-time Fourier transform. Cheng et al. [17] used a unique model based on continuous wavelet transform and local binary convolutional neural networks (CWT-LBCNN); the input vibration signal adaptively extracts features, enabling automatic fault diagnosis of the rotating machinery (RM). Unlike traditional CNNs, the CWT-LBCNN replaces the conventional convolutional layer with a local binary convolution layer, allowing for faster training and more excellent resistance to overfitting. Liang et al. [18] presented Wavelet Transform, Generative Adversarial Networks, and convolutional neural network (WT-GAN-CNN) for rotating machinery, a unique and highly accurate fault detection technique that integrates CNNs, (GANs), and (WT). The three key components of the (WT-GAN-CNN) technique is outlined. To begin, WT takes one-dimensional raw time-domain signals and uses them to extract features of images based on time and frequency. Second, more training picture examples are produced by GANs. Ultimately, the constructed CNN model is employed to detect faults in rotating machinery using the generated real and false training time-frequency images. Altaf et al. [19] used a single microphone to record the audible sound of the machine being tested, and various statistical, spectral, and Spectro-temporal data were then extracted. A variety of machine learning approaches, that is, classifiers like K-nearest neighbor (KNN), support vector machine (SVM), kernel linear discriminant analysis (KLDA), and sparse discriminant analysis (SDA), are then used to assess the chosen features. Based on simulation results, defects such as ball faults, inner race, and outer race defects it is classified. Amini et al. [20] utilized (AE) envelope analysis to identify faults in train axle bearings. To make a comparison, they recorded two types of datasets: one from freight wagons with intentionally defective axle bearings and the other from a custom-built test rig in a laboratory setting. Li et al. [21] applied (AE) technology, combined with vibration signals, SVM, and wavelet packet decomposition, to detect and classify faults in aero-engine bearings. The method first used wavelet packet decomposition to break down the original vibration signal into various frequency bands. Next, the energy values of each frequency band were extracted as feature vectors. Finally, these feature vectors were input into an SVM for fault detection and classification. Bahedh et al. [22] introduced a revolutionary two-dimensional color-coded map termed the normalized-diagnostic-feature-gram (NDFgram) alongside a deep (ANN) to create a diagnostic method that is independent of rotational speed. They utilized this technique for bearing flaw analysis to evaluate its efficacy. Specifically, First, acquired bi-frequency spectral coherence (SCoh) data, and second, computed diagnostic features (DFs) by synthesizing the SCoh data throughout the spectral frequency domain utilizing the center and range of frequencies. Subsequently, they illustrated the DFs on a two-dimensional map, positioning them relative to the center and frequency resolution, and employed stacked maps for diverse defect features to formulate the analytical models. Finally, the characteristic model was trained, and bearing defects were identified using a pre-trained (CNN). The efficacy of the proposed approach was compared with other methods that use different input characteristics for CNN. The results showed that the proposed method achieved a 100% success rate and 98.16% accuracy for constant-speed tests, with a defect analysis accuracy of 98.56% for the second type of ball bearing. Overall, the proposed methodology outperformed others in terms of diagnostic accuracy. Najim and Alsalaet [23] utilized envelope analysis characteristics as inputs for (ANN) to tackle a cross-domain learning issue. addressed three principal cross-domain challenges, types of roller bearings, rotational velocities, and loading circumstances, by deriving supplementary characteristics at fundamental defect frequencies from the original signal through the envelope method. They employed three varieties of roller bearings, three rotational velocities, and three fault classifications to replicate all cross-domain jobs experimentally. The data for a certain bearing type under designated operating conditions were utilized to train the ANN, while data from other bearing types were employed to assess the method's efficacy. Under stable load conditions, the average success rate for Koyo bearings was recorded at 99.5%, the average accuracy for NU bearings at 98.33%, and the average accuracy for the faulty bearing kit at 97.3%

## 2. Theoretical background

### *2.1. Kurtogram*

The Kurtogram is a powerful fourth-order spectral analysis tool used to generalize SK for a given vibration signal $x(n)$. The SK method identifies a series of transients and their impacts in the frequency domain by selecting an appropriate frequency band. For each frequency component of $x(n)$, the corresponding SK value is calculated to characterize the transients of $x(n)$ as a function of frequency [7].

The kurtosis coefficient has been widely applied for detecting bearing faults and is derived from time-domain data. The variance of a zero-mean signal (a signal without a DC component) is the mean square value, equivalent to the square of the root-mean-square (RMS) value. AC-coupled accelerometer signals are inherently zero-mean. However, a rectified waveform has a nonzero average value. In such cases, the variance is expressed as:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x(t_i)-\mu)^2 \quad (1)$$

Where, $N$ is the number of data points in the sequence, and $\mu$ represents the average value.

The kurtosis coefficient is a fourth-order statistical measure normalized by the square of the variance and is calculated as:

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\frac{(x(t_i)-\mu)^4}{\sigma^4} \quad (2)$$

High values of $\gamma$ indicate significant deviations from the RMS value [24].

*2.1.1. Mathematical derivation of the Kurtogram*

To obtain the Kurtogram, follow these steps:

1. Spectral Kurtosis (SK) Definition

Spectral Kurtosis SK(f) is defined as:

$$SK(f) = \frac{E[|X(f,t)|^4]}{(E[|X(f,t)|^2])^2} - 2 \quad (3)$$

Where:

$X(f, t)$ is the Short-Time Fourier Transform (STFT) of the signal $x(t)$, $E[\cdot]$ denotes expectation (statistical mean), $|X(f, t)|^4$ and $|X(f, t)|^2$ represent the fourth and second power of the magnitude spectrum, respectively [25].

2. Filtering and the Wavelet Packet Transform (WPT)

- The signal is decomposed into multiple frequency bands using a wavelet packet decomposition or Finite Impulse Response (FIR) filter banks.
- The decomposition follows a dyadic tree structure, where each level refines the frequency resolution by splitting bands further [26].

3. Kurtosis Computation for Each Band

- The time-domain signal from each frequency band is extracted.
- The kurtosis of each filtered signal is computed as:

$$K_x = \frac{E[(x-\mu)^4]}{\sigma^4} \quad (4)$$

where: $\mu$ is the mean of $x$, $\sigma$ is the standard deviation of $x$.

- The optimal frequency band is the one with the maximum kurtosis value, indicating the presence of impulsive (fault-induced) features [27].

A kurtogram is a visual representation of a signal that illustrates the energy distribution across various frequency levels. This analysis is typically performed using methods such as wavelet transform or spectrum analysis. In Matlab, the kurtogram is generated using a specialized kurtogram function, which analyzes the signal.

The kurtogram function requires the input signal segment and the segment length (number of samples) to produce the visual output. The resulting graph displays the energy density, or kurtosis, across different frequency bands.

Figures 1 and 2 show the samples of the obtained diagnostic patterns for the Koyo 1205 bearing for acoustic emission and vibration data by the kurtogram with (a) normal, (b) outer race, (c) inner race, and (d) ball defect. At speed 15 Hz, 25 Hz, 30 Hz, and 35 Hz respectively.

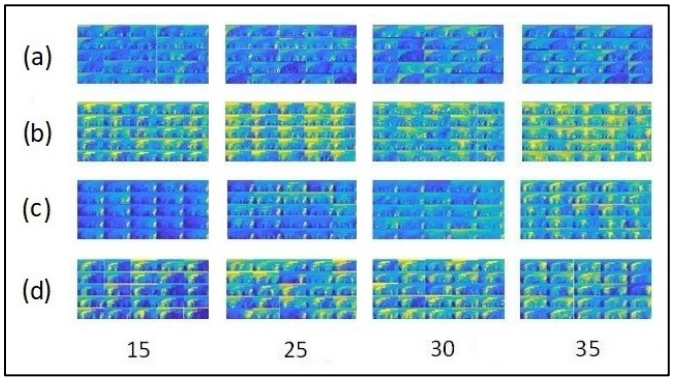



**Fig. 1** Outputs kurtogram of state (AE).

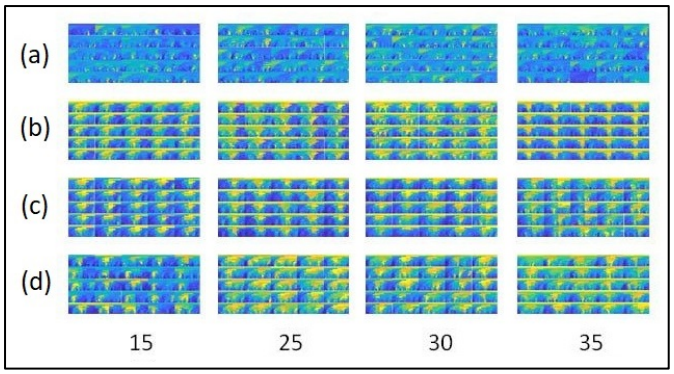



**Fig. 2** Outputs kurtogram of state vibration.

*2.2. A Deep learning neural network*

The structure of a Deep Learning Neural Network is summarized in this section, with more detailed explanations available in a Deep Learning Neural Network is a multi-stage neural network composed of three primary stages: the filter stage, the convolutional layer, and the pooling layer. These stages are designed to extract features from input data. Additionally, the classification stage, implemented as a multi-layer perceptron, is typically used for final prediction. Each layer in the network has specific functionalities, which are detailed below.

*2.2.1. Convolutional layer*

The convolutional layer processes input regions by applying filter kernels, followed by an activation unit to produce output features. Each filter kernel extracts specific local features from input regions, effectively capturing spatial hierarchies. In this layer, $K^l$ and $b^l$ denote the weights and biases of the $l$-th filter kernel, and $x_j^l(t)$ represents the $j$-th input region in the $l$-th frame. The convolution operation is expressed as:

$$y_j^l(t) = K^l x_j^l(t) + b^l \tag{5}$$

*2.2.2. Activation layer*

After convolution, the activation function is critical in introducing non-linearity, enhancing feature representation, and improving learning capabilities. The Rectified Linear Unit (ReLU) is widely used because it accelerates convergence and simplifies backpropagation adjustments. The formula for ReLU is:

$$y_j^l(t) = \max\left(0, f\left(y_j^l(t)\right)\right) \tag{6}$$

Where, $f(y_j^l(t))$ is the convolution output, and $y_j^l(t)$ represents the activation result.

*2.2.3. Pooling layer*

The pooling layer is essential for reducing the spatial size of feature maps, which decreases the computational requirements and enhances invariance to input transformations. A common operation is max pooling, defined as:

$$p_i^{l+1}(t) = \max_{i \in [j - W + 1, j + W]} q_i^l(t) \tag{7}$$

Here, $q_i^l(t)$ refers to the value of the $i$-th neuron, $W$ is the width of the pooling region, and $p_i^{(l+1)}(t)$ represents the neuron value after pooling.

*2.2.4. Batch normalization*

Batch Normalization (BN) is a technique that minimizes internal covariance shifts, thereby accelerating training. Typically placed after the convolutional or fully connected layer and before activation, (BN) normalizes the input $x = x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ as:

$$\hat{x}^{(i)} = \frac{x^{(i)} - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{8}$$

$$y^{(i)} = \gamma\, \hat{x}^{(i)} + \beta \tag{9}$$

Where, $\mu$ and $\sigma^2$ represent the mean and variance of the input, $\gamma$ and $\beta$ are learnable scale and shift parameters, and $y^{(i)}$ is the normalized output. Batch Normalization ensures stability and improves network performance by reintroducing learned scales [5].

*2.2.5. AlexNet*

AlexNet is a sophisticated (CNN) for image processing and classification. It utilizes interconnected layers to extract features from input images and make classification decisions. Therefore, in this study, structured initially with five convolutional layers and two fully connected layers, AlexNet was designed for categorizing images into 1000 classes based on the ImageNet dataset. The final architecture is shown in Fig. 3. Also, the CNN architecture is implemented in Table 1. The network has been partially modified for the current classification task, which involves four classes.

The original input size for AlexNet is 227 × 227 × 3 (RGB channels). Since the RGB images in this task have dimensions 875 × 656 × 3, they are resized to fit the required input size. Preprocessing through zero-centered normalization is applied to reduce data variance and enhance performance.

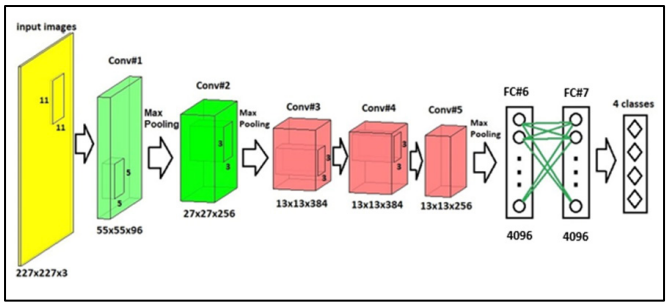



**Fig. 3** The architecture of the modified AlexNet was used for classification in this research.

The network begins with the first convolutional layer, which has 96 filters of size 11 × 11, a stride of 4, and no padding. This reduces the image dimensions to 55 × 55 × 96 and extracts basic features like edges and patterns. A ReLU activation function introduces non-linearity by converting negative values to zero, accelerating the learning process. Cross-channel normalization is applied to reduce inter-channel variance and improve training stability. Dimensionality is further reduced using max pooling with a 3 × 3 filter and a stride of 2, simplifying data while preserving essential features. The second convolutional layer consists of two groups of 128 filters, each sized 5 × 5 with 2 × 2 padding. This layer extracts more complex features, like corners and intricate patterns. Similar operations (convolution, ReLU, normalization, and pooling) are repeated across the subsequent convolutional layers. The third, fourth, and fifth convolutional layers contain 384, 384, and 256 filters, respectively, with smaller 3 × 3 filters and strides to capture finer details. Another pooling operation 3 × 3 in the fifth layer reduces dimensions to 6 × 6 × 256, preparing the data for the fully connected layers. The network then includes three fully connected layers, each with 4096 nodes. These layers convert spatial data into a linear representation and integrate the extracted features. Dropout is used during training to randomly disable some connections, reducing overfitting and enhancing generalization. The Softmax layer converts the output into probabilities corresponding to the classification categories. The final output has dimensions 1 × 4, with each value representing the probability of a particular class. The class with the highest probability is chosen as the final prediction. This structured network processes tasks sequentially, starting with basic feature extraction (e.g., edges), advancing to complex features (e.g., structures), and culminating in classification.

*2.3. Analysis parameters and methodology for signal processing using the Kurtogram*

In signal analysis using the Kurtogram, several parameters play a critical role in identifying and analyzing distinct patterns within a signal.

The Kurtosis Coefficient serves as a measure of the sharpness or peakedness of data, indicating the presence of impulsive signals or sudden changes, such as faults. By calculating Kurtosis across various time-frequency bands, regions with the most distinct patterns can be pinpointed.

**Table 1.** The architecture of the applied CNN.

| Type | Dimensions | Details |
|---|---|---|
| Input data (RGB image) | 227 × 227 × 3 | Zero-centered normalization |
| Convolutional | 55 × 55 × 96 | 96 11 × 11 × 3 filters, stride of [4 4], and zero-padding of 0, 0, 0, 0 |
| ReLU | 55 × 55 × 96 | |
| Cross-channel normalization | 55 × 55 × 96 | |
| Max pooling | 27 × 27 × 96 | 3 × 3 max pooling, stride of [2 2], and padding 0, 0, 0, 0 |
| Convolutional | 27 × 27 × 256 | 2 groups of 128 5 × 5 × 48 filters, stride of [1 1], and zero-padding of 2, 2, 2, 2 |
| ReLU | 27 × 27 × 256 | |
| Cross-channel normalization | 27 × 27 × 256 | |
| Max pooling | 13 × 13 × 256 | 3 × 3 max pooling, stride of [2 2], and padding 0, 0, 0, 0 |
| Convolutional | 13 × 13 × 384 | 384 3 × 3 × 256 filters, stride of [1 1], and zero-padding of 1, 1, 1, 1 |
| ReLU | 13 × 13 × 384 | |
| Convolutional | 13 × 13 × 384 | 2 groups of 192 3 × 3 × 192 filters, stride of [1 1], and zero-padding of 1, 1, 1, 1 |
| ReLU | 13 × 13 × 384 | |
| Convolutional | 13 × 13 × 256 | 2 groups of 128 3 × 3 × 192 filters, stride of [1 1], and zero-padding of 1, 1, 1, 1 |
| ReLU | 13 × 13 × 256 | |
| Max pooling | 6 × 6 × 256 | 3 × 3 max pooling, stride of [2 2], and padding 0, 0, 0, 0 |
| Fully connected | 1 × 1 × 4096 | |
| ReLU | 1 × 1 × 4096 | |
| Dropout | 1 × 1 × 4096 | |
| Fully connected | 1 × 1 × 4096 | |
| ReLU | 1 × 1 × 4096 | |
| Dropout | 1 × 1 × 4096 | |
| Fully connected | 1 × 1 × 4 | |
| Softmax | 1 × 1 × 4 | |
| Classification output | 1 × 1 × 4 | 4 classes in the output |

The process involves dividing the signal into multiple analysis levels, each further subdivided into smaller frequency bands. The optimal level is determined by identifying the maximum value of Kurtosis, denoted as $K_{max}$.

Another key parameter is the window length, which refers to the duration of the signal segment under examination. The accuracy of the analysis depends on the window length: shorter windows yield higher temporal resolution, while longer windows offer greater frequency resolution. The ideal window length is selected based on analysis outcomes.

Bandwidth represents the range of frequencies being analyzed. The signal is split into distinct frequency bands, and Kurtosis is computed for each band to identify the one with the most significant information. The center frequency of this optimal band is then determined and used to design a filter that focuses on the most meaningful frequencies.

To ensure accuracy, noise removal is a crucial step. The Kurtogram helps isolate bands containing valuable signal information while minimizing noise interference. Additionally, the type of window function, such as Hanning or Hamming, is chosen to enhance spectral analysis and reduce distortions from window edges.

Steps for signal analysis using Kurtogram:

1. Signal decomposition: the signal is broken down into various analysis levels using methods like filter banks or wavelet packet decomposition.
2. Kurtosis calculation: Kurtosis is calculated for each time-frequency band, and the maximum value $K_{max}$ is identified.
3. Optimal level selection: the level corresponding to the highest $K_{max}$ is selected for further analysis.
4. Bandwidth and center frequency determination: the bandwidth and center frequency of the band with the most distinctive signal are established.
5. Filtering: a filter is designed to target the optimal bandwidth, extracting the meaningful signal while reducing noise.
6. Analysis of the filtered signal: the filtered signal is examined to detect faults or extract useful features.

This systematic approach allows for precise fault detection and feature extraction by focusing on the most significant time-frequency components of the signal. The complete diagnostic approach can be summarized in Fig. 4 block diagram of the approach.

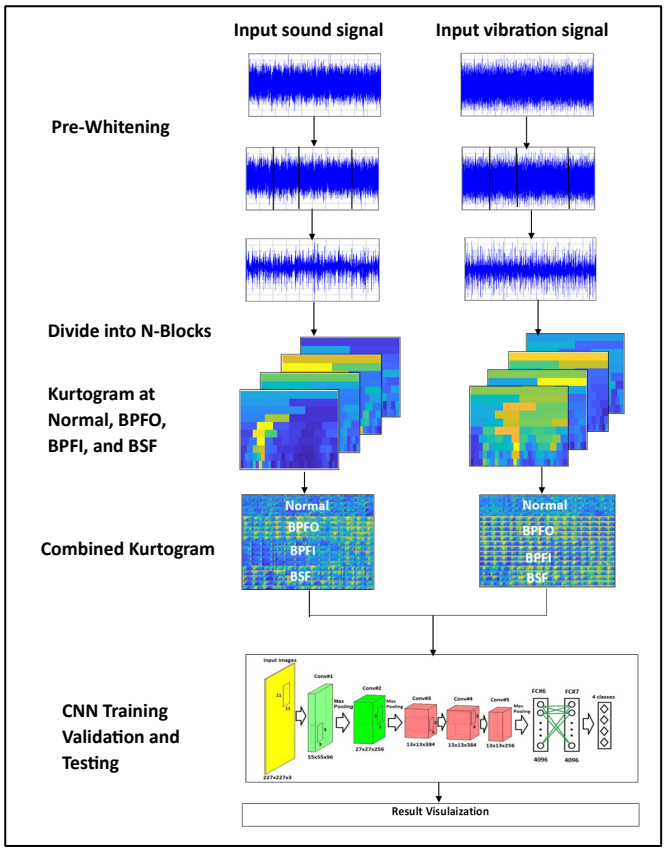



**Fig. 4** Block diagram of the proposed diagnosis approach.

## 3. Experimental setup

### *3.1. Description of machinery fault simulator (MFS)*

It is a device used in an effective and direct way for the purpose of examining and diagnosing faults in rotating machines. It was manufactured with complete precision to ensure smooth diagnosis without any harmful interference to the received signals, as shown in Fig. 5 available at the University of Basrah, College of Engineering, which consists of a variable frequency alternating electric motor with a power of one horsepower and a frequency of 60 Hz, accompanied by an AC motor control unit that allows changing the speed. To measure the rotational speed, a tachometer equipped with an LCD screen is used. The device also includes steel shafts with diameters of 3/4 inch and 1 inch, with the shaft size adjusted to suit the dimensions of the bearing being studied. A disc is fixed in the middle of the shaft by two clamping rings. During the test, a normal bearing is placed in the coupling side housing (flexible coupling) and a test bearing with certain errors on the other side. A, B, and K 4366 accelerometer, serial number 0931214, with an industrial microphone model 426E01 is installed on the test bearing side to capture the sound and vibration signals together and send them to the IDAC-245 data acquisition device, which collects the vibration and sound data and sends them to the computer, where the data is then analyzed to obtain information. The main components used in generating the sound and vibration signals as shown in Fig. 5 include the following:

1. Electric motor.
2. Tachometer.
3. AC control unit.
4. Flexible coupling.
5. Normal bearing.
6. Test bearing.
7. 5 kg disc placed in the middle of the shaft.
8. Industrial microphone to measure the sound signals.
9. Accelerometer to capture the vibration signals.
10. Data acquisition device IDAC-245.

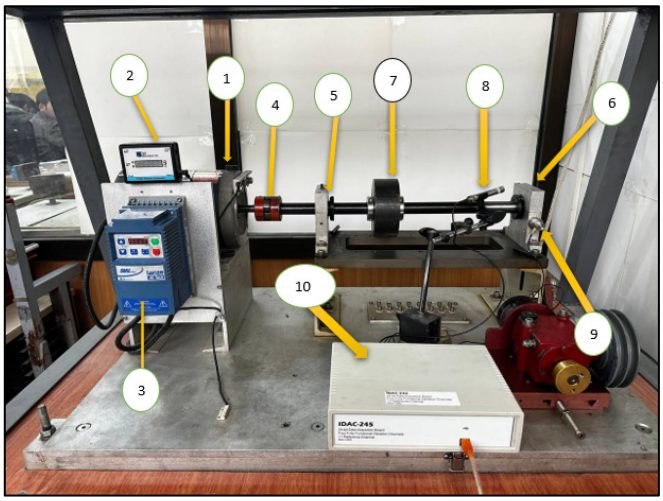



**Fig. 5** MFS components.

Used one good and three faulted bearings (BSF, BPFO, BPFI).

The 378B02 industrial microphone in Fig. 6 from PCB Piezotronics is a half-inch pre-polarized condenser microphone tailored for acoustic measurements in industrial and research settings. Equipped with a built-in preamplifier and Transducer Electronic Data Sheet (TEDS) technology, it facilitates seamless automatic identification and calibration when paired with compatible devices. Renowned for its precision and reliability, the 378B02 is widely used in industrial acoustic measurements, research experiments, and environmental noise analysis applications. The physical characteristics of this microphone are given in Table 2 [28].


**Fig. 6** industrial microphone.

**Table 2.** Physical properties of microphone model 378B02.

| 378B02 Prepolarized Free-Field Microphone System | | |
|---|---|---|
| Normal Microphone diameter | in (mm) | 1/2 (12) |
| Sensitivity at 250 Hz (± 1.5 dB) | mV/Pa (dB re 1 V/Pa) | 50 (-26) |
| Frequency range (± 2 dB) | Hz | 3.75 – 20000 |
| Frequency range (± 1 dB) | Hz | 7 – 10000 |
| Cartridge thermal noise (Microphone) | dB[A] re 20 µPa | 15 |
| Inherent noise with 426E01 preamp | dB[A] re 20 µPa | 15.5 |
| Harmonic distortion limit: 3% | dB re 20 µPa | 147 |
| Distortion limit with 426E01 preamp | dB re 20 µPa | 137 |

### *3.2. The double-row self-aligning ball bearing*

A double-row self-aligning ball bearing is a type of rolling element bearing designed to accommodate misalignment and shaft deflection, making it well-suited for applications

involving high vibration or misalignment. In this study, the self-aligning bearing model 1205, manufactured by Koyo, is deemed critical due to its role in facilitating more precise and comprehensive experiments. This bearing's significance is further emphasized as datasets collected at 1200 rpm serve as the training data for all proposed diagnostic techniques, with fault detection relying on classifiers trained on these datasets. Additionally, the 1205 bearing offers ease of assembly and disassembly, even when parts sustain damage. Figure 7 illustrate the bearing and its individual components.

**Fig. 7** Shows the Koyo 1205 bearing in its assembled state.

The manufacturer calculated bearing fault frequencies for 3/4" bore rotor bearings. Bearing fault frequencies, also referred to as bearing characteristic frequencies, are determined by the dimensions and specifications of the bearing. These frequencies can be obtained from the manufacturer or calculated using bearing equations. Figure 8 shows the cross-section of the self-aligning ball bearing, while Table 3 provides the dimensions of the Koyo 1205 bearing along with the calculated bearing frequencies.

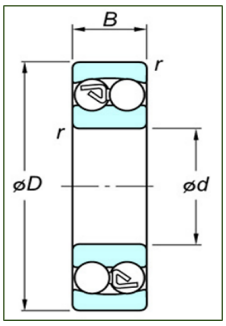


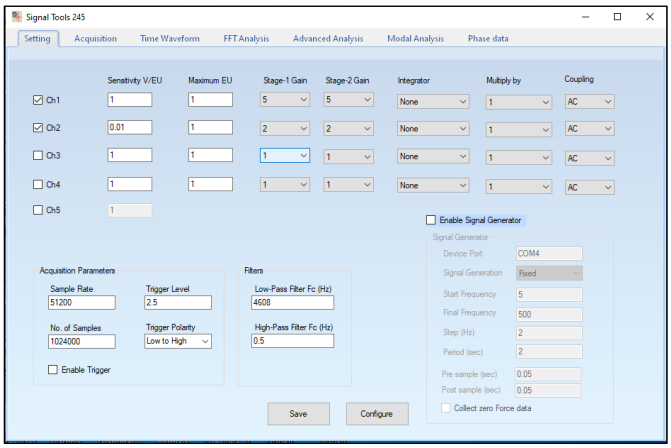

**Fig. 8** Cross-sectional view of the self-aligning ball bearing.

**Table 3.** Koyo 1205 bearing dimensions and calculated fault frequencies.

| Dimensions | Value | Unit |
|---|---|---|
| B, Width | 15 | mm |
| D, Outer diameter | 52 | mm |
| d, Inner diameter | 25 | mm |
| Ball diameter | 7 | mm |
| Number of balls per row | 12 | - |
| Defect location | Fault frequency per revolution | |
| BPFO | 4.891 | |
| BPFI | 7.11 | |
| BSF | 2.57 | |

*3.3. Experiment procedure*

1. Select the experimental bearing sample that contains defects and install the experimental bearing in the fault simulation device.
2. Determine the required rotation speed in the test.
3. Start the engine after ensuring all safety requirements and ensure that all installation tools are safe.
4. Read the sound and vibration wave from the sound sensors (industrial microphone) and vibration (accelerometer) for a period estimated at 10 seconds via the (IDAC-245) device used to read the waves.
5. Save the read data in the computer and turn off the power to the engine to end the test.
6. Analyze the data using the signal tool program see Fig. 9, where convert the waves obtained during the test from the time domain to the frequency domain.




7. After converting the data to the frequency domain, we use the envelope analysis to ensure the appearance of the frequencies required for the type of defect in the bearing.
8. After ensuring that all the required frequencies appear, use the SK to convert the saved data into two-dimensional color images that are later programmed using the Matlab software.
9. After saving the two-dimensional color images for each type of defect, whether inner, outer, or compound, they are used in the CNN program to classify each type of defect, where 70% of the total color images are used in the training process, while 30% of the total images are used for CNN authentication or checking the CNN results.

## 4. Results and discussion

In this paper, the theoretical and experimental results will be discussed and explained. The results will divide into two parts based on the data acquisition method. The first part of the results will include the results obtained from the sound sensor, where four types of bearings were used, three defective and one healthy, with four different speeds, which are 15 Hz, 25 Hz, 30 Hz, and 35 Hz, respectively. Also, 1600 images were used in the CNN for classification purposes, where 70% of the images were used for training purposes and 30% for verification purposes, as shown in Fig. 10, where the training took about 118 minutes and 36 seconds, and the number of complete training cycles was 6 Epochs, the total number of repetitions was 672, the number of repetitions in each cycle was 112, and the maximum number of repetitions reached was 672, where the verification process was done every 3 repetitions. The model reached an accuracy of 92.50% on the validation data, indicating good model performance. As for training performance, the graph shows that the training accuracy starts very low (around 20%) in epoch 1 and reaches around 95-98% as training progresses. This indicates that the model is learning patterns from the training data efficiently, reflecting a good network structure and an appropriate learning rate. The loss starts at a very high value (more than 4 in epoch

1), but quickly drops to below 0.1 in epoch 6. The sharp drop at the beginning is normal, as the model tries to adjust its initial weights. As for validation performance, the validation accuracy improves similarly to the training accuracy, reaching 92.5%, which is an excellent value compared to most models. The relative match between the training and validation curves indicates that the model does not suffer from problems such as overfitting or underfitting. As for the validation loss, it drops very quickly at the beginning, but slows down in improvement as the cycles progress. In the end, the loss is very low, indicating good convergence. As for the comparison between training and validation, there are minor differences as there is not much difference between the loss and accuracy of training and validation, which means that the model does not just memorize the training data, but generalizes its performance to new data. This indicates that the balance between the number of layers, the number of nodes, and other parameters is well-tuned. Stability in the last epochs in epochs 5 and 6, observe stability in performance (consistency in accuracy and loss), indicating that the model has reached the final learning stage and there are no more significant improvements. As for the statistical analysis of the final performance, the final validation accuracy of 92.5% means that 92.5% of the predictions on the validation data were correct and is high enough to be used in most applications. This is very good, especially if the domain in which the model is applied is completely insensitive to errors. The low final loss value indicates that the model is very good at predicting the correct values, as the difference between the predictions and the actual results is small. The variance in training accuracy compared to validation accuracy shows a very small difference, which means that the model does not suffer from the problem of overfitting.

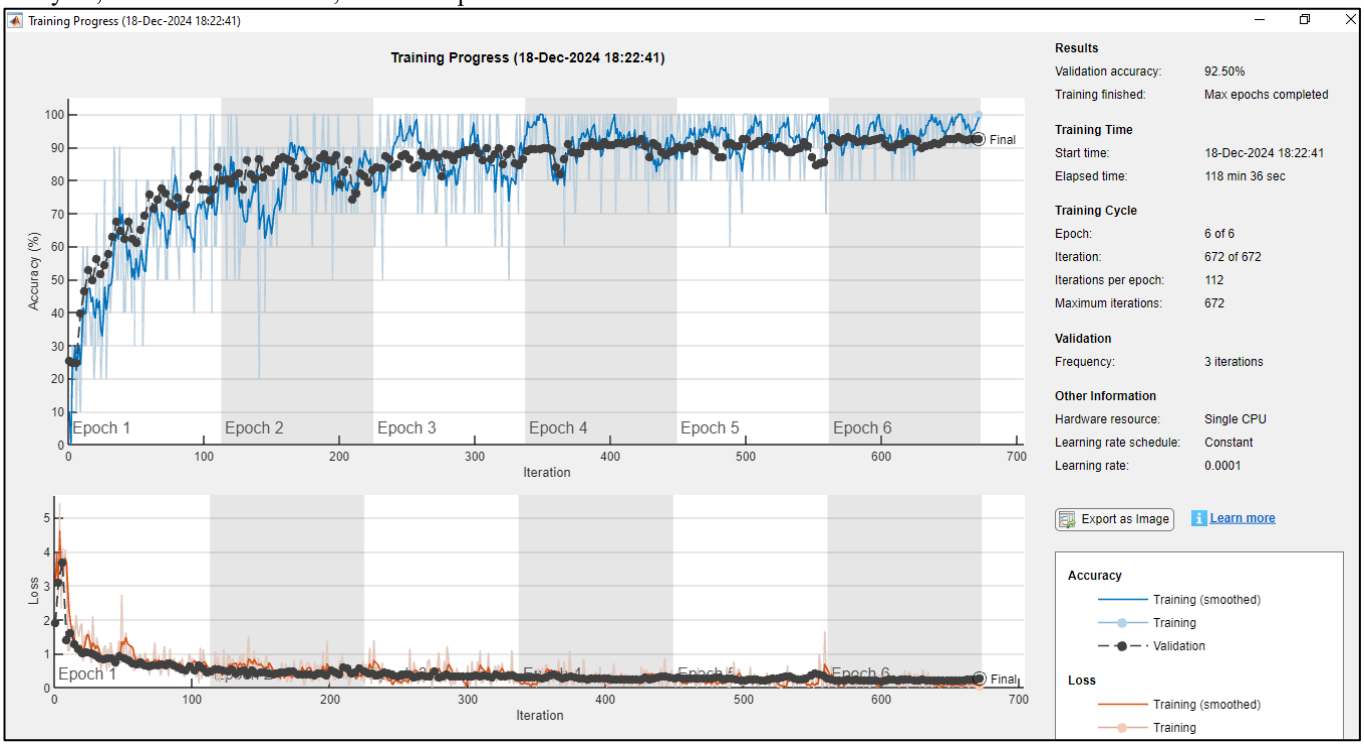



**Fig. 10** Training progress results for sound data at four different speeds, namely 15 Hz, 25 Hz, 30 Hz, and 35 Hz, respectively.

A confusion matrix in Fig. 11 is derived from analyzing an acoustic signal for diagnosing rotating machinery faults. The matrix illustrates the model's performance in classifying four operational conditions: ball, inner, normal, and outer. The classification accuracy is measured based on the number of correctly classified samples versus misclassifications.

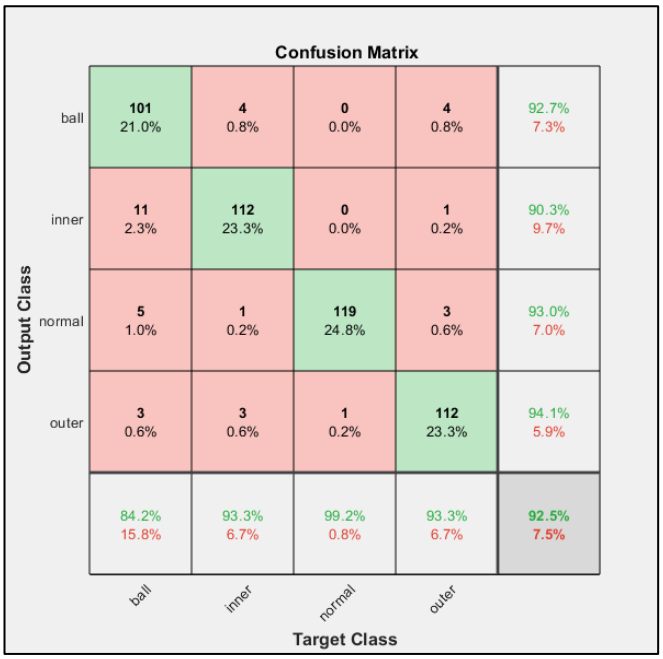



**Fig. 11** Confusion matrix.

The relative values in each matrix cell reflect the model's precision and ability to distinguish between different fault types accurately, notice that in the first row, the total number of samples for the ball category was 109 samples, where the classification accuracy for this category was 92.7%, i.e. 101 of the samples were classified correctly, and 7.3%, i.e. 8 of the samples were classified incorrectly, as follows: 4 of the samples, i.e. 0.8%, were incorrectly classified as being from the inner category, and 0.8%, i.e. 4 of the samples were classified as being from the outer category. As for the general targeted accuracy for this category, it was 84.2%, and the error rate was 15.8%. As for the second class, we note that the total number of samples for the inner category was 124 samples. The accuracy for this inner category was 90.3%, i.e. 112 samples were classified correctly, and the error rate was 9.7%, i.e. 12 of the samples were classified incorrectly, and 2.3%, i.e. 11 of the samples were incorrectly classified as a ball, and 0.2%, i.e. 1 of the samples were incorrectly classified as outer. As for the general targeted accuracy for this category, it was 93.3%, and the error rate was 6.7%. As for the third class, notice that the total samples for the normal category were 128 samples, where the accuracy for this category was 93%, i.e. 119 samples were classified correctly, and the error rate was 7%, i.e. 9 samples were classified incorrectly, as follows: 5 samples, i.e. 1%, were incorrectly classified as ball, 0.2%, i.e. 1 of the samples were incorrectly classified as inner, and 0.6%, i.e. 3 of the samples were incorrectly classified as outer. As for the general targeted accuracy for this normal category, it was 99.2%, and the error rate was 0.8%. As for the fourth class, notice that the total number of outer samples is 119 samples, where the accuracy for this class was 94.1%, meaning 112 of the samples were correctly classified, and the error rate in the classification was 5.9%, meaning 7 of the samples were incorrectly classified, as follows: 0.6%, meaning 3 of the samples were incorrectly classified as ball, 0.6%, meaning 3 of the samples were incorrectly classified as inner, and 0.2%, meaning 1 of the samples were incorrectly classified as normal. As for the targeted accuracy for this class, it was 93.3%, and the error rate was 6.7%. From the matrix, we notice that the total accuracy of the model is 92.5%, and the error rate is 7.5%. The model is generally accurate, and the performance on all classes is very good. The highest accuracy was for the 99.2% normal class, and the lowest accuracy was for the 84.2% ball class.

The second part explain for the results obtained from the vibration sensor, where four types of bearings were used, three defective and one healthy, with four different speeds, which are 15 Hz, 25 Hz, 30 Hz, 35 Hz, respectively, and 1600 images were used in the CNN for classification purposes, where 70% of the images were used for training purposes and 30% for verification purposes, as shown in Fig. 12, where the training took about 2 hours and 27 seconds, and the number of complete training cycles was 6 epochs, the total number of iterations was 672, the number of iterations in each cycle was 112, and the maximum number of iterations reached was 672, where the verification process was done every 3 iterations. The model achieved a very high accuracy 94.38% with a low and stable loss. The model's performance on the training data (blue line) and the verification data (black dots) is very close, which is a positive indicator because the model does not suffer from overfitting or underfitting. The simple oscillation in the validation accuracy, it is shown that the data is not completely identical between training and validation, which is normal, and means that the model handles the complexity of the data well. The loss decreases rapidly in the beginning epoch 1 and 2, indicating that the model quickly learned the basic patterns in the data. As we progress in later epochs from epoch 4 to 6, the loss decrease slows down, indicating that the model is approaching saturation i.e. learning everything that can be learned from the current data. In the end, the model performs excellently and its indicators are positive in terms of accuracy and loss. It can be considered ready for practical application based on the high validation accuracy.

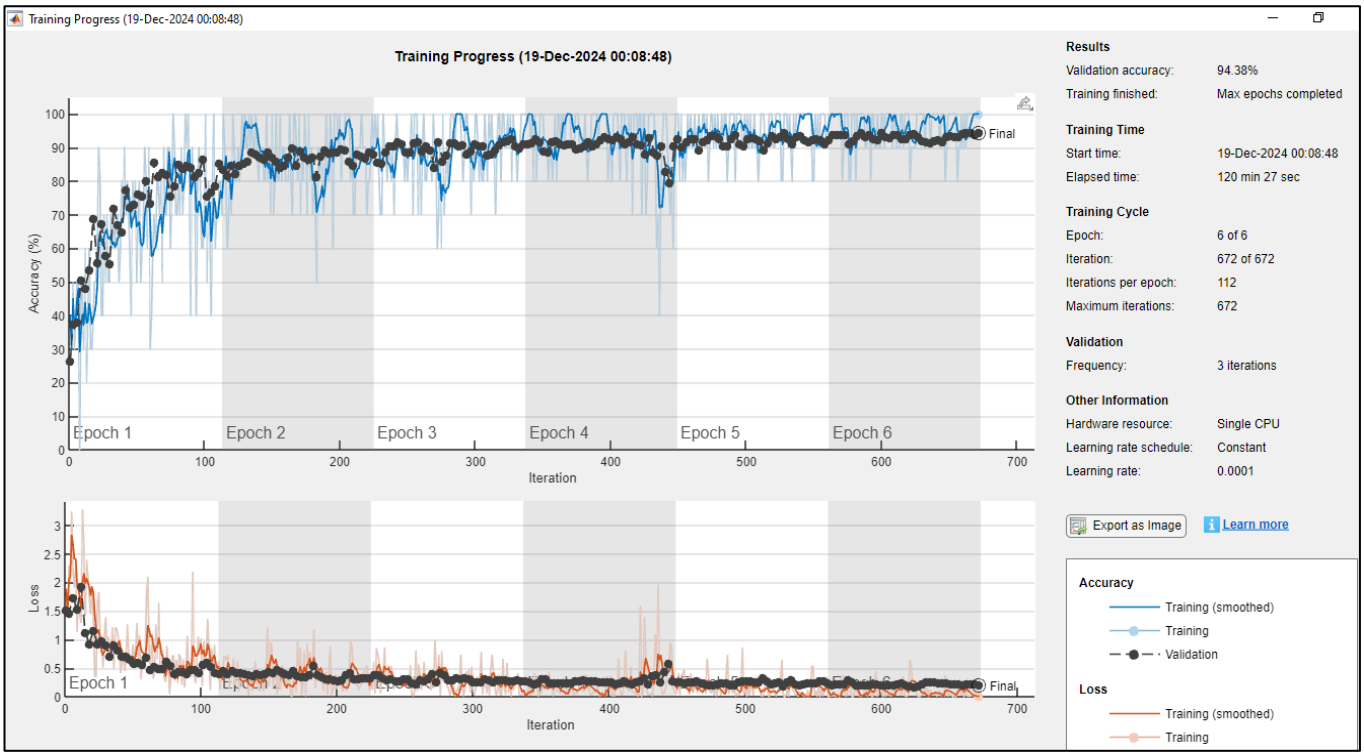



**Fig. 12** Training progress results for vibration data at four different speeds, namely 15 Hz, 25 Hz, 30 Hz, and 35 Hz, respectively.

Figure 13 shows a confusion matrix is derived from analyzing a vibration signal to diagnose rotating machinery faults. The matrix illustrates the model's performance in classifying four operational conditions: ball, inner, normal, and outer. Classification accuracy is measured based on the number of correctly classified samples versus misclassifications. The relative values in each matrix cell reflect the model's precision and ability to distinguish between different fault types accurately. From the first-row notice that the model's accuracy in predicting the ball is 90.8%, meaning that 90.8% of the predictions in which the model said the ball were correct, while the percentage of the sample that was classified among all the actual samples of the same category is 90%. Also, for the inner category, the model's prediction accuracy for the inner category is 95.7%, while the actual percentage among all the samples of the same category is 93.3%. As for the normal category, the model's prediction accuracy is 94.5%, while the actual percentage among all samples is 100%, meaning that the model never made a mistake in identifying the normal category.

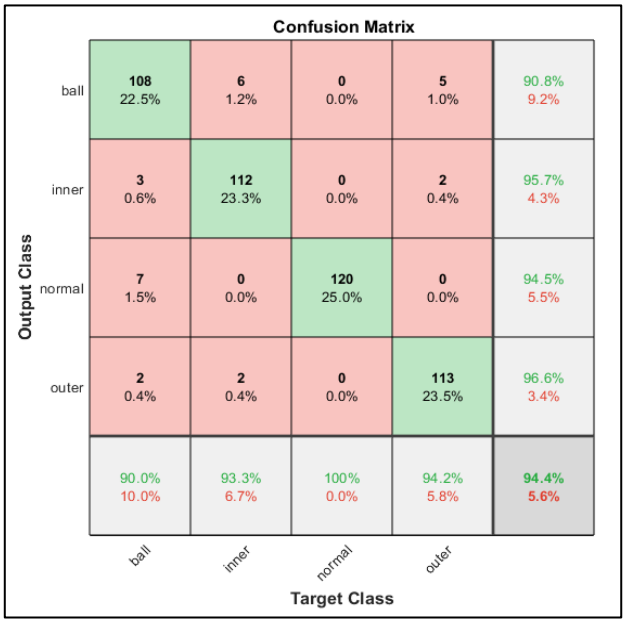



**Fig. 13** Confusion matrix.

As for the outer category, the model's prediction accuracy was 96.6%, while the actual percentage among all samples was 94.2%. In general, the overall accuracy of the model is 94.4%. This indicates that the model is generally strong, especially in classifying categories such as outer and normal. Errors are present, but they are limited and concentrated among categories with great similarity such as ball and inner.

## 5. Conclusions

The proposed approach involves gathering vibration and sound data under both healthy and faulty bearing conditions, extracting relevant features, and training a model to identify various bearing states. The findings demonstrate that using

CNN with vibration data achieves notably higher classification accuracy. The models are tested on four bearing conditions three faulty and one healthy and CNN combined with data from both sensors consistently delivers superior performance in terms of classification accuracy, precision, and other metrics. The results highlight that leveraging vibration and sound signatures with CNN allows for early and accurate detection of bearing issues. This capability is critically important in the industry, as it can save lives and reduce the high costs associated with machine failures caused by cumulative bearing damage. Consequently, the demand for such diagnostic tools has surged significantly in recent times.

## References


[1] M. M. M. Islam, and J.-M. Kim, "Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network," Computers in Industry, Vol. 106, pp. 142-153, 2019. https://doi.org/10.1016/j.compind.2019.01.008

[2] W. Qian, S. Li, J. Wang, Z. An and X. Jiang, "An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network," Measurement Science and Technology, Vol. 29, Issue 9, 2018. https://doi.org/10.1088/1361-6501/aad101

[3] R. Huang, Y. Liao, S. Zhang and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," IEEE Access, Vol. 7, pp. 1848-1858, 2019. https://doi.org/10.1109/ACCESS.2018.2886343

[4] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," IEEE Transactions on Instrumentation and Measurement, Vol. 67, Issue 1, pp. 185-195, 2017. https://doi.org/10.1109/TIM.2017.2759418

[5] 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, Issue 2, 2017. https://doi.org/10.3390/s17020425

[6] S. S. Udmale, S. S. Patil, V. M. Phalle and S. K. Singh, "A bearing vibration data analysis based on spectral kurtosis and ConvNet," Soft Computing, Vol. 23, Issue 19, pp. 9341-9359, 2019. https://doi.org/10.1007/s00500-018-3644-5

[7] S. S. Udmale, S. K. Singh, and S. G. Bhirud, "A bearing data analysis based on kurtogram and deep learning sequence models," Measurement, Vol. 145, pp. 665-677, 2019. https://doi.org/10.1016/j.measurement.2019.05.039

[8] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," Mechanical systems and signal processing, Vol. 100, pp. 439-453, 2018. https://doi.org/10.1016/j.ymssp.2017.06.022

[9] N. Günnemann, and J. Pfeffer, "Predicting defective engines using convolutional neural networks on temporal vibration signals," Proceedings of Machine Learning Research, Vol. 74, pp. 92-102, 2017.

[10] D.-T. Hoang, and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," Cognitive Systems Research, Vol. 53, pp. 42-50, 2019. https://doi.org/10.1016/j.cogsys.2018.03.002

[11] H. Wang, S. Li, L. Song, L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," Computers in Industry, Vol. 105, pp. 182-190, 2019. https://doi.org/10.1016/j.compind.2018.12.013

[12] W. Zhang, G. Peng, and C. Li, "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input," MATEC web of conferences, Vol. 95, 2017. https://doi.org/10.1051/matecconf/20179513001

[13] Z. Chen, A. Mauricio, W. Li, K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," Mechanical Systems and Signal Processing, Vol. 140, 2020. https://doi.org/10.1016/j.ymssp.2020.106683

[14] Y. Cao, Y. Ding, M. Jia, R. Tian, "A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings," Reliability Engineering & System Safety, Vol. 215, 2021. https://doi.org/10.1016/j.ress.2021.107813

[15] S. Xiong, H. Zhou, S. He, L. Zhang, Q. Xia, J. Xuan, and T. Shi, "A novel end-to-end fault diagnosis approach for rolling bearings by integrating wavelet packet transform into convolutional neural network structures," Sensors, Vol. 20, Issue 17, 2020. https://doi.org/10.3390/s20174965

[16] H.-Y. Chen, and C.-H. Lee, "Deep learning approach for vibration signals applications," Sensors, Vol. 21, Issue 11, 2021. https://doi.org/10.3390/s21113929

[17] Y. Cheng, M. Lin, J. Wu, H. Zhu, X. Shao, "Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network," Knowledge-Based Systems, Vol. 216, 2021. https://doi.org/10.1016/j.knosys.2021.106796

[18] P. Liang, C. Deng, J. Wu, Z. Yang, "Intelligent fault diagnosis of rotating machinery via wavelet transform, generative adversarial nets and convolutional neural network," Measurement, Vol. 159, 2020. https://doi.org/10.1016/j.measurement.2020.107768

[19] M. Altaf, M. Uzair, M. Naeem, A. Ahmad, S. Badshah, J. A. Shah and A. Anjum, "Automatic and efficient fault detection in rotating machinery using sound signals," Acoustics Australia, Vol. 47, pp. 125-139, 2019. https://doi.org/10.1007/s40857-019-00153-6

[20] A. Amini, M. Entezami, and M. Papaelias, "Onboard detection of railway axle bearing defects using envelope analysis of high frequency acoustic emission signals," Case Studies in Nondestructive Testing and Evaluation, Vol. 6, pare A, pp. 8-16, 2016. https://doi.org/10.1016/j.csndt.2016.06.002

[21] H. Li, Z. Wu, K. Xue and G. Yang, "Research on aero-engine bearing fault using acoustic emission technique based on wavelet packet decomposition and support vector machine," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, pp. 1444-1448, 2017. https://doi.org/10.1109/IAEAC.2017.8054252

[22] J. K. Alsalaet, A. Hajnayeb, and A. S. Bahedh, "Bearing fault diagnosis using normalized diagnostic feature-gram and convolutional neural network," Measurement Science and Technology, Vol. 34, Issue 4, 2023. https://doi.org/10.1088/1361-6501/acad1f

[23] H. S. Najim, and J. K. Alsalaet, "Cross-domain diagnosis of roller bearing faults based on the envelope analysis adaptive features and artificial neural networks," Journal of

Vibration and Control, Vol. 30, Issue 13-14, pp. 3168-3183, 2024. https://doi.org/10.1177/10775463231191684

[24] S. A. McInerny, and Y. Dai, "Basic vibration signal processing for bearing fault detection," IEEE Transactions on education, Vol. 46, Issue 1, pp. 149-156, 2003. https://doi.org/10.1109/TE.2002.808234

[25] J. Antoni, "The spectral kurtosis: a useful tool for characterising non-stationary signals," Mechanical systems and signal processing, Vol. 20, Issue 2, pp. 282-307, 2006. https://doi.org/10.1016/j.ymssp.2004.09.001

[26] Y. Yan, F. Gao, and F. F. Yap, "Numerical model of spindle/disks assembly–shaft–housing system for vibro-acoustic analysis of HDD in idle mode," Mechanical systems and signal processing, Vol. 20, Issue 2, pp. 438-462, 2006. https://doi.org/10.1016/j.ymssp.2005.01.004

[27] T. Barszcz, and R. B. Randall, "Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine," Mechanical systems and signal processing, Vol. 23, Issue 4, pp. 1352-1365, 2009. https://doi.org/10.1016/j.ymssp.2008.07.019

[28] PCB Piezotronics, Model 378B02 Prepolarized Condenser Microphone. https://www.pcb.com/products?m=378b02