The 4th International Conference on Electrical Machines and Drives (ICEMD 2024), 24-25 December, 2024, Tehran, Iran.

Acoustic Fault Diagnosis of Rolling Bearings in Induction Motors Using Time-Frequency Image Analysis

Erfan Madadi Mahani
department of Railway Engineering
Iran University of Science and Technology
Tehran, Iran
erfan_madadimahani@rail.iust.ac.ir

Ahmad Mirabadi
department of Railway Engineering
Iran University of Science and Technology
Tehran, Iran
mirabadi@iust.ac.ir

Ali Rezazadeh
department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
ali_rezazadeh@elec.iust.ac.ir

Amir Hossein Karamali
department of Electrical Engineering
Iran University of Science and Technology
Tehran, Iran
a_karamali@elec.iust.ac.ir

***Abstract*— Detecting bearing failures in induction motors is essential for preventing unexpected breakdowns and minimizing industrial downtime. The evaluation of acoustic signals and their spectral characteristics provides a non-invasive method for identifying bearing faults. Over the years, various monitoring techniques have been introduced, incorporating both sound and vibration sensors. Among these, rolling element bearing analysis has become a well-established approach for monitoring rotating machinery conditions. However, the implementation of sophisticated accelerometer-based bearing monitoring systems, which are widely adopted in the industry, may not be cost-effective for non-critical machinery due to extended return on investment periods. Consequently, human auditory perception often serves as an initial diagnostic tool for interpreting bearing noise. This study examines the effectiveness of machine learning and deep learning techniques in diagnosing rolling element bearings. It further explores a classification method that converts 1-D bearing sound signals into 2-D scalogram images and utilizes transfer learning with a pre-trained model. This approach substantially reduces the time required for feature extraction and selection in traditional bearing diagnostics while preserving high accuracy for the dataset employed in this research.**

Keywords—non-invasive fault detection, condition monitoring, machine learning, deep learning, scalogram, transfer learning.

I. Introduction

Condition-based maintenance and rotational fault diagnosis systems play a crucial role in modern industries. Monitoring the lifespan of machinery is essential for timely maintenance. Key industrial objectives include reducing maintenance costs while improving safety, efficiency, reliability, and availability of critical equipment. Around 30%–40% of motor failures result from rolling bearing defects, typically appearing as surface damage on components such as the outer race, inner race, cage, or rollers [1].

Recently, the field of acoustic analysis has garnered increasing interest and has found application across various domains. One such example is the limited utilization of speech recognition in industrial settings for condition monitoring[2].

This is predominantly attributed to the fact that investing in acoustic systems may lack practicality in the presence of cost-effective accelerometer-based systems. Generally, accelerometer-based systems are widely embraced in various industries as a norm due to their reduced susceptibility to background noise compared to microphone-based systems. Nevertheless, in scenarios involving non-critical and relatively inexpensive uses, the expense associated with implementing a specialized diagnostic system could exceed the value of the component being monitored, such as in the context of low voltage electric motors[3].

As shown in [4] vibration analysis is a highly sensitive and reliable technique for detecting non-stationary bearing fault patterns. Similarly, sound signal analysis has been recognized as an effective method for diagnosing bearing faults, as it operates without physical contact, allowing access to hazardous areas. Additionally, unlike vibration sensors, it does not require complex mounting configurations and remains a cost-effective solution.Nevertheless, the sound signals are significantly attenuated by machinery or outside noise, as evidenced by the authors in [4] which presents a challenge in discerning the disturbance related time series signals from the background noise. To enhance diagnostic efficacy in the context of variable speed motors or background noise, several authors have proposed the integration of vibration and sound signals as a means of improving diagnostic performance [4],[5].

In recent years, the predominant approach to diagnosing bearing health has centered on integrating time-frequency-based methods and applying artificial intelligence, such as machine learning (ML) or deep learning (DL) algorithms, to

979-8-3315-0751-0/24/$31.00 ©2024 IEEE

analyze the feature patterns. Continuous wavelet transform (CWT) stands out as a highly efficient technique for transforming bearing time-series signals into feature images[6],[7].

In high-frequency applications, analyzing non-stationary signals with the standard CWT involves a localized function that lacks translation invariance, which can introduce instability. A scalogram provides a visual representation of a signal in the time-frequency domain, utilizing a two-dimensional image to depict both time and frequency components[8]. Pre-trained networks have undergone training on more than a million images, enabling them to develop comprehensive feature representations. Transfer learning is a widely used technique in deep learning applications. In this study, we fine-tune pre-trained convolutional neural networks to classify scalograms. Recently, deep learning architectures have shown significant advantages over ML classifiers, primarily due to their ability to automatically extract features[9].

Some studies have shown that the Convolutional Neural Network (CNN) stands out as a highly important and effective DL model, featuring sparse connectivity and shared weights, particularly when analyzing images related to bearing health, whether in colormap or grayscale. In comparison to conventional ML models, CNNs eliminate the need for manual feature extraction or a distinct denoising method[10]. The use of 2-D CNN, along with feature images, has a notable impact on enhancing the accuracy of classification[11]. In the context of classification tasks involving complex data sets or non-stationary backgrounds, recent research has proposed an intelligent fusion or hybrid scheme combining a conventional ML algorithm with DL model [12]. Prior research has highlighted the effectiveness of bearing fault diagnosis with high accuracy. However, implementing feature extraction algorithms based on complex signal processing techniques can be time-intensive, computationally demanding, and may necessitate additional resources such as cost, power, and infrastructure for real-time industrial use. Furthermore, the challenges posed by harsh industrial environments must be considered alongside resource constraints. Achieving precise diagnostic performance with a single sound sensor remains difficult, especially under varying operational conditions and external disturbances[4],[13]. It was observed by researchers that the artificial intelligence-based (Wavelet Scattering Transform) WST-aided extracted features demonstrated superior performance in the detection of bearing faults[14]. In this approach, WST-based features are extracted from current signals and fed into multiple ensemble learning models, including random forest (RF) and extreme gradient boosting (XGB), alongside an artificial neural network (ANN). This hybrid method demonstrates strong potential for handling variations in speed and load conditions [14]. Additionally, the authors [8]indicated that statistical features, derived from the vibration-based WST network, were fed to the 1-D-CNN under fixed speed and load conditions. Nevertheless, it is essential to investigate the impact of external machinery disturbances on the measurement when the actuator under examination operates in an industrial setting with fluctuating operating conditions. The hybrid CNN and support vector machine (SVM) model [13]demonstrate a shorter test time (TT) than the hybrid CNN and fuzzy c-means clustering (FCM) approach [15]. Furthermore, the combined CNN-SVM method has been demonstrated to achieve higher classification accuracy in remote sensing applications and in biomedical applications[16]. However, this method has not yet been independently tested using either vibration or sound signals alone. While previous intelligent models have largely aimed at improving performance through architectural enhancements, the current priority lies in assembling comprehensive and representative datasets for training and evaluation. The quality of these datasets will ultimately play a crucial role in determining the model's effectiveness in real-world industrial settings. In[17], a smart fault diagnosis approach utilizing the Cross Wavelet Transform (XWT) and the GoogleNet model is introduced. This method applies XWT and bandpass filtering to analyze acoustic signals from a traction motor bearing fault test bench at two different positions. Initially, XWT is employed to identify the frequency band associated with the fault feature. A bandpass filter is then applied to eliminate noise-related frequency components, retaining only the fault-related frequencies. Finally, the kurtosis spectrum of both the filtered and original signals is fed into GoogleNet for error classification. The results indicate that GoogleNet achieves 98.23% accuracy in classifying errors using filtered signals, whereas its accuracy drops to 89.66% when classifying original signals.In[18], a hybrid FN-CNN-SVM approach combined with Wavelet Scattering Transform (WST) was proposed for analyzing motor vibration and sound signals under simulated challenging industrial conditions. The experimental setup involved three primary conditions: (1) fluctuations in speed, (2) variations in load, and (3) an unstable working environment. Data was gathered from laboratory-based process control systems, monitored through a supervisory control and data acquisition framework. The proposed model demonstrated exceptional accuracy, reaching 100% in fault detection based on sound and vibration signals.

II. Data Description

This article uses the open-access dataset ”Vibration, Acoustic, Temperature, and Motor Current Dataset of Rotating Machine Under Varying Load Conditions for Fault Diagnosis”[15]. The rotating machine was tested under three distinct conditions: normal operation, inner race bearing defects, and outer race bearing defects. To capture both vibration and acoustic signals, the Siemens SCADAS Mobile 5PM50 system was utilized. The dataset comprises 60-second recordings for the normal state, along with an equivalent duration for each fault scenario. In order to minimize interference from air-cooled brakes, acoustic data were recorded at a 51.2 kHz sampling rate exclusively under no-load conditions.

III. Metodology

In this section, we compare various ML and DL methods and introduce our proposed approach to enhance fault diagnosis accuracy.

To assess the performance of our models, accuracy was utilized as the evaluation metric. Accuracy quantifies the overall correctness of the model's predictions and is calculated using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

where TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False Negative when classifying the validation data into the three classes[8].

A. Machine learning methods

In this section, K-nearest neighbor (KNN), decision tree (DT), RF, and SVM are used for bearing health classification. We augment the dataset by dividing it into 0.1 seconds long data, and Various traditional statistical features like standard deviation (STD), kurtosis, skewness, root mean square (RMS), and maximum amplitude are extracted. The Formula for the mentioned statistical features are as follows:

$$STD=\sqrt{\frac{\sum_{i=1}^{N}[x(i)-\mu]^2}{N}} \quad (2)$$

$$Kurtosis = \frac{\sum_{i=1}^{N}\left[x(i)-\mu\right]^4}{N \times \sigma^4} \quad (3)$$

$$Skewness = \frac{\sum_{i=1}^{N}\left[x(i)-\mu\right]^3}{N \times \sigma^3} \quad (4)$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[x(i)\right]^2} \quad (5)$$

where $x(i)$, μ , N ,and σ are the i-th data point in the data samples, the mean (average) of the data samples, the total number of data points in the data samples, and standard deviation respectively[8].

Table 1 shows that RF has the highest accuracy among statistical feature-based classifiers (87%).

TABLE I
MACHINE-LEARNING CLASSIFICATION PERFORMANCE

Classifier	Accuracy (%)
SVM	82.77
DT	86.44
RF	**87.00**
KNN	86.22

B. Long Short-Term Memory (LSTM) networks

LSTM networks, a variant of the RNN architecture, have become widely recognized for their effectiveness in handling long-term dependencies within sequential data. Equipped with memory cells, LSTMs can retain information over extended periods, allowing them to utilize contextual details across an entire input sequence efficiently. This feature makes LSTMs especially effective for handling sequential data, including applications like speech recognition, audio classification, and time series analysis. The architecture of LSTM networks comprises memory cells with gating mechanisms, including forget and input gates, which regulate the flow of information within the network[19]. These gating mechanisms allow LSTMs to control the retention or removal of information at each time step, aiding in the learning of long-range dependencies while addressing the vanishing and exploding gradient issues often seen in conventional RNNs. In our approach, 0.1-second segments of audio signals, corresponding to different bearing conditions, were processed by the LSTM network, resulting in an accuracy of 98.88%.

C. Scalogram

Scalograms visually represent the wavelet transform (WT). Unlike sinusoidal functions, WTs use a wavelet basis to create linear time-frequency representations. By incorporating a scale variable in addition to the time variable, the WT is well-suited for analyzing nonstationary and transient signals. For a wavelet transform, WT_x, of a signal which is energy limited $x(t) \in L^2\left(R\right)$; the basis for the transform can be set as

$$WT_x(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(t)\,\psi(\frac{t-b}{a})dt \quad (6)$$

where a is the scale parameter, b is the time parameter, and ψ is analyzing wavelet.

Various studies have explored the effectiveness of different wavelets in matching a signal. Available options include Gaussian, Morlet, Shannon,Laplace, Hermit, and Mexican Hat wavelets, applicable in both simple and complex functions. At present, there is no universally accepted approach for selecting the optimal wavelet, making it an ongoing subject of debate in research[16]. The Morlet wavelet, $\psi\sigma(t)$, is selected due to its effectiveness in handling sign errors commonly found in mechanical systems and is mathematically expressed as:

$$\psi_\sigma(t) = c_\sigma \pi^{-(\frac{1}{4})}\, e^{-(\frac{1}{2})t^2}(e^{i\sigma t} - K_\sigma) \quad (7)$$

where c is normalization constant, and $K\sigma$ is admissibility criterion[20].

A scalogram represents the absolute value of the continuous wavelet transform (CWT) of a signal, visualized in terms of time and frequency. It can be more effective than a spectrogram when analyzing real-world signals that contain multiple scale features, such as gradual variations disrupted by abrupt changes. In contrast, a spectrogram is generated by applying a fixed-length window to the input signal, which is then shifted across time and frequency. This window remains uniform, real-valued, and non-oscillatory, leading to the formation of the spectrogram.

The spectrogram has a fixed time-frequency resolution due to its constant window. On the other hand, the CWT is achieved by applying a scaled and time-shifted wavelet to the signal. This wavelet can be complex-valued and oscillates. The prototype wavelet undergoes scaling and shifting operations in the CWT. Scaling both shrinks and stretches the wavelet, creating short-term high frequency wavelets when reduced in size, and long-lasting, low-frequency wavelets when stretched.

These wavelets are particularly effective for identifying transient events and distinguishing prolonged, low-frequency occurrences. The process of computing a scalogram involves the following steps:

• If the signal contains more than one million samples, divide it into overlapping segments.

• Apply the CWT to each segment to generate the corresponding scalogram.

• Present the scalogram sequentially, segment by segment..

D. Scalogram of Bearing Data

The presence of distinct peaks is a key feature for differentiating between inner race defects, outer race defects, and normal conditions. Consequently, scalograms present a promising approach for classifying bearing faults. In this study, all bearing signal measurements were collected from experiments conducted at a consistent shaft speed. To effectively use signals from different shaft speeds, the data must be standardized accordingly; otherwise, the number of "pillars" in the scalogram may not be accurate. The scalogram highlights unique peaks corresponding to impulses in the audio signal. The 1-D sound signals were converted into scalograms and stored as images for training. Each scalogram initially had dimensions of 227×227×3, later resized to 224×224×3 to align with the input size requirements of pre-trained networks. Finally, all scalogram images related to the bearing signals were stored, and an image datastore was created. The dataset was then divided, with 80% of the images used for training and the remaining 20% allocated for validation to assess model performance..

E. Pre-trained image classification models

Pre-trained image classification models serve as highly effective tools for computer vision applications, particularly in image recognition and classification. These models undergo training on extensive datasets containing labeled images, enabling them to identify intricate patterns and correlations between visual features and their respective object categories.

The training process involves feeding millions of images into the model, each labeled with the corresponding object class. The model analyzes these images, extracts features, and learns to associate these features with the correct class labels. As the model processes more images, it refines its ability to generalize and accurately classify new images. The training process involves feeding millions of images into the model, each labeled with the corresponding object class. The model analyzes these images, extracts features, and learns to associate these features with the correct class labels. As the model processes more images, it refines its ability to generalize and accurately classify new images.

The benefits of using pre-trained image classification models include reduced training time and resources, improved accuracy, transfer learning, and accessibility.

F. Training the Network Using Transfer Learning

Next, modify the existing convolutional neural network to perform classification on the scalograms.

Transfer learning is a widely employed technique in deep learning scenarios. It involves leveraging a pre-trained network as a foundation for a fresh task. The process of fine-tuning a network through transfer learning is typically more efficient and less complex compared to training a network with weights initialized randomly. This approach allows for the swift transfer of acquired features using a reduced set of training images. Fig.1 illustrates the operational flow of the proposed bearing fault detection system.Within this framework, bearing sound signals are first recorded and preprocessed. The 1-D audio signals are then transformed into 2-D scalograms, capturing their time-frequency characteristics. These scalogram images are fed into a pre-trained deep learning model, such as ResNet or MobileNet, to classify bearing conditions into three categories: normal, inner race fault, and outer race fault. By utilizing transfer learning, this method enhances training efficiency and ensures high accuracy in fault detection.

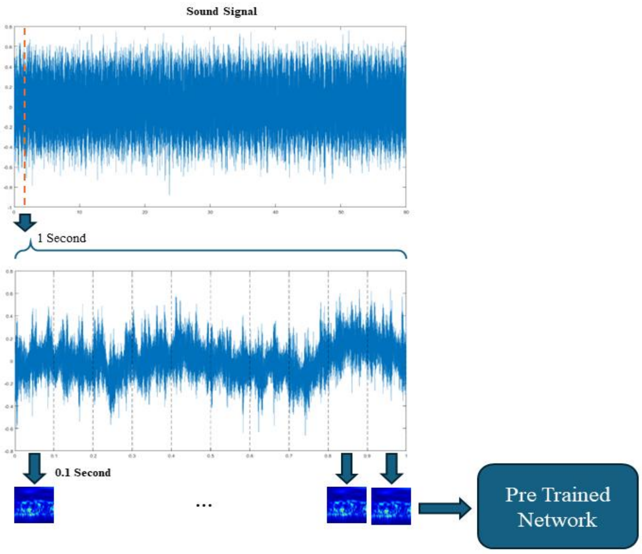

Fig. 1. Functional diagram of the proposed system

G. Validate Using Test Data Sets

Evaluate the trained network's accuracy using bearing signal test data. The test data must undergo identical preprocessing steps as the training data. Utilize the trained model to categorize images within the test datastore and analyze the precision of its predictions.

IV. EVALUATION OF RESULT

This section delves deeper into the experimental details, including the dataset and evaluation methods. This study introduces an approach for acoustic diagnosis of rolling bearings utilizing deep learning models. The process involves segmenting the data into 0.1-second intervals, generating scalograms, and subsequently inputting them into pre-trained image classification models such as SqueezeNet, ResNet, and MobileNet. Fig. 2 illustrates the time-frequency representation of bearing sound signals across three different conditions: (a) normal operation, (b) inner race defect, and (c) outer race defect. Within these scalograms, specific frequency and intensity peaks indicate the bearing's condition in each scenario. Notably, the inner and outer race faults exhibit unique peak distributions that highlight structural anomalies in the bearing. These distinctive characteristics enable the deep learning model to effectively differentiate between various bearing states.

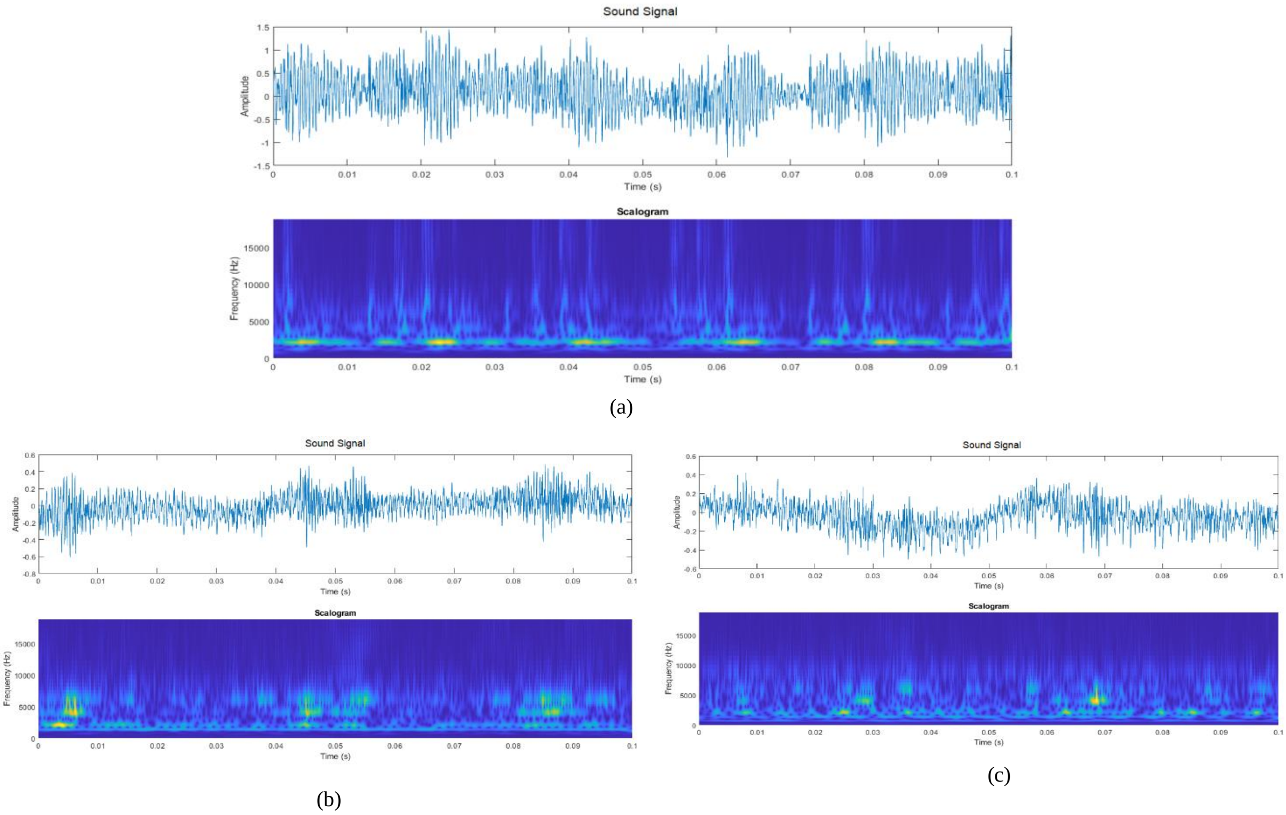

Fig. 2. Scalogram of Bearing Data (a) Normal operation (b) Inner Race defect (c) Outer Race defect.

TABLE II
CNN CLASSIFICATION PERFORMANCE

Model	Accuracy (%)
Squeeznet	99.63
Resnet-50	**100**
MobileNet	**100**

A. Discussion

The findings of this study indicate that employing scalograms as input for deep learning models is an effective approach for classifying rolling bearing faults. This method has demonstrated high accuracy across various image classification tasks and is applicable to multiple domains. One key advantage is its simplicity, allowing for seamless adaptation to different datasets and models. Additionally, it benefits from the availability of well-trained, pre-trained deep learning models. However, some challenges remain. A major limitation is the need for a large dataset to properly train the models. Furthermore, the computational cost can be high, particularly for complex deep learning architectures. Despite these challenges, this approach remains a reliable tool for diagnosing rolling bearing faults.

V. CONCLUSION

In this article, we have leveraged the available dataset for diagnosing bearing failures using artificial intelligence algorithms. We segmented the audio signal into 0.1-second intervals to augment the data. We then extracted various statistical features such as standard deviation, kurtosis, and others from the signal and fed them into machine learning algorithms for classification. It was observed that the random forest algorithm achieved the highest accuracy (87%). Next, we fed the audio signals into an LSTM neural network and the accuracy improved to 98.88%. Finally, we presented our method in which scalograms are generated from the audio signal, and the images are fed to pre-trained image classification models. It was observed that we achieved 100% accuracy with this method using ResNet and MobileNet network.

REFERENCES

[1] O. Yaman, "An automated faults classification method based on binary pattern and neighborhood component analysis using induction motor," *Measurement*, vol. 168, p. 108323, 2021.

[2] C.-H. Lee, J.-S. Jwo, H.-Y. Hsieh, and C.-S. Lin, "An intelligent system for grinding wheel condition monitoring based on machining sound and deep learning," *IEEE Access*, vol. 8, pp. 58279-58289, 2020.

[3] T. Mian, A. Choudhary, and S. Fatima, "An efficient diagnosis approach for bearing faults using sound quality metrics," *Applied Acoustics*, vol. 195, p. 108839, 2022.

[4] G. Wu, N. Yan, K.-n. Choi, H. Jung, and K. Cao, "A two-step vibration-sound signal fusion method for weak fault feature detection in rolling bearing systems," *Advances in Mechanical Engineering*, vol. 13, no. 12, p. 16878140211067155, 2021.

[5] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network," *Measurement,* vol. 173, p. 108518, 2021.
[6] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, no. 6, pp. 2658-2669, 2019.
[7] M.-Q. Tran, M.-K. Liu, Q.-V. Tran, and T.-K. Nguyen, "Effective fault diagnosis based on wavelet and convolutional attention neural network for induction motors," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-13, 2021.
[8] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock and Vibration,* vol. 2017, no. 1, p. 5067651, 2017.
[9] M. R. Bhuiyan and J. Uddin, "Deep transfer learning models for industrial fault diagnosis using vibration and acoustic sensors data: A review," *Vibration,* vol. 6, no. 1, pp. 218-238, 2023.
[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.
[11] J. Wang, D. Wang, S. Wang, W. Li, and K. Song, "Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network," *IEEE Access,* vol. 9, pp. 23717-23725, 2021.
[12] D. Zhang, Y. Chen, F. Guo, H. R. Karimi, H. Dong, and Q. Xuan, "A new interpretable learning method for fault diagnosis of rolling bearings," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-10, 2020.
[13] A. Biswas, S. Ray, D. Dey, and S. Munshi, "Detection of simultaneous bearing faults fusing cross correlation with multikernel SVM," *IEEE Sensors Journal,* vol. 23, no. 13, pp. 14418-14427, 2023.
[14] R. N. Toma *et al.*, "Classification framework of the bearing faults of an induction motor using wavelet scattering transform-based features," *Sensors,* vol. 22, no. 22, p. 8958, 2022.
[15] W. Jung, S.-H. Kim, S.-H. Yun, J. Bae, and Y.-H. Park, "Vibration, acoustic, temperature, and motor current dataset of rotating machine under varying operating conditions for fault diagnosis," *Data in brief,* vol. 48, p. 109049, 2023.
[16] Z. Feng, M. Liang, and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples," *Mechanical systems and signal Processing,* vol. 38, no. 1, pp. 165-205, 2013.
[17] G. Yang, Y. Wei, and H. Li, "Acoustic Diagnosis of Rolling Bearings Fault of CR400 EMU Traction Motor Based on XWT and GoogleNet," *Shock and Vibration,* vol. 2022, no. 1, p. 2360067, 2022.
[18] S. Mitra and C. Koley, "Real-time robust bearing fault detection using scattergram-driven hybrid CNN-SVM," *Electrical Engineering,* vol. 106, no. 3, pp. 3615-3625, 2024.
[19] A. Karamali, A. Daeichian, and S. Rezaei, "Using Long Short-Term Memory Networks as Virtual Wind Direction Sensors for Improved Wind Farm Turbines Orientation," in *2024 9th International Conference on Technology and Energy Management (ICTEM)*, 2024: IEEE, pp. 1-5.
[20] J. Lin and L. Qu, "Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis," *Journal of sound and vibration,* vol. 234, no. 1, pp. 135-148, 2000.