Paper ID: 416

DIAGNOSIS AND PREDICTION OF BEARING FAULT BY DEEP LEARNING APPROACH WITH ENVELOPE ANALYSIS

Raihan Karal[1], Md. Emdadul Hoque[1], Shah Mohammed Takrim[1], Md. Farhan Shakib[2]

[1]Department of Mechanical Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
[2]Department of Computer Science and Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
E-mail: raihan.karal@me.ruet.ac.bd

Abstract

Predictive maintenance for machinery aims to anticipate malfunctions before they occur. Bearings in rotating machinery are especially prone to wear, misalignment, and other forms of degradation. With advances in data collection under the Industry 4.0 framework, particularly in gathering vibration data, various methods have emerged to predict deterioration more accurately. In this work, a widely-used dataset from Case Western reserve University (CWRU), which includes detailed information on bearing failures, is utilized. Vibration analysis is applied to this dataset to extract key features that can be used to develop an optimized deep learning model for predictive maintenance. A model is proposed based on artificial neural Network (ANN) to predict and diagnosis bearing fault. The bearing vibration data from CWRU at different working condition were extracted into different statistical features. Synthetic minority over-sampling technique (SMOTE) was applied to address the limited amount of data in this dataset. The features were standardized using Standard Scaler. The model employed in this work is a feedforward ANN designed for a classification task. The network consists of an input layer, three hidden layers, and an output layer. Moreover, the outcomes of these assessments demonstrate that this model functions effectively when compared to other models. This accuracy achieved of this model is 98% with very low processing time (Training time less than 1 second per epoch). Considering all the things, this model can be applied in the actual world and is dependable, appropriate, and capable of identifying bearing defects under various operating situations.

Keywords: ANN, Machine learning, *SMOTE*, Vibrational accelerometer

1. Introduction

Machines are very essential part of the industries. There are numerous variety of machines in the industries. The profit of an industry directly depends on proper operation and running of machine. To inflate the profit margin of an industry the machine elements are required to run properly and smoothly. Improper maintenance strategy can enhance maintenance cost of the industry significantly. Rotating elements of machine are very essential part of machines. Bearings are necessary for the smooth operation of rotating equipment. But misalignment, wear, and lubrication problem can degrade bearings over time [1]. Different fault diagnosis methods classify and detect these problems. Several model-driven techniques rely on analyzing vibration, temperature, and wear debris. Among these, vibration analysis is particularly valuable as it can provide extensive insights into potential anomalies[2]. Traditional data-driven fault diagnostic models typically employ machine learning algorithms, classifiers, and signal processing methods to detect issues.

Durjay et al. [1], utilizing a dataset from CWRU, developed a model that effectively combines transfer learning with the VGG16 model for feature extraction and random forest for classification, demonstrating strong performance. Additionally, a neural network-based early fault detection system has been created for small and medium-sized induction motors, leveraging commonly available data such as rotor speed and stator current. Real-time evaluation on a 3/4 hp permanent magnet motor shows satisfactory accuracy of over 95% [3]. Y.S. Wang et al. [4] proposed a noise-based HHT-SVM model for engine fault diagnosis, utilizing Hilbert-Huang transform and

support vector machine techniques. The model measures engine noises, extracts fault features, and trains an optimal support vector machine (SVM) model. The method is accurate and effective, suitable for both stationary and nonstationary signals. D.H. Pandya et al. [5] proposed a fault diagnosis technique using acoustic emission analysis with Hilbert-Huang Transform (HHT) to analyze intrinsic mode functions (IMFs). The K-nearest neighbor (KNN) and APF-KNN algorithms were used for efficient classification, demonstrating improved reliability in diagnosing rolling element bearing faults. Miao H. et al. [6] introduces a deep learning-based method for bearing fault diagnosis, utilizing a large memory storage retrieval (LAMSTAR) neural network and short-time Fourier transform. Validated using a bearing test rig, the method shows accurate classification performance, even at low rotating speeds, ensuring safety and normal operation. The SVM was used as the primary algorithm, with particle swarm optimization (PSO) technique improving performance. The model outperformed traditional ML techniques like KNN, DT, and LDA in classification accuracy, with improved accuracy of 93.9% [2].

Alonso-gonzález et al. [7] uses vibration analysis and envelope analysis on the CWRU bearing dataset to detect faults with minimal data and features. By optimizing bandwidth with a kurtogram, accuracy improved, achieving 94.4% with Kernel Naive Bayes, and a perfect 100% with Decision Tree and KNN models. These results show effective fault detection with streamlined feature selection. Raj et al. [8] uses raw accelerometer data and neural networks for bearing fault classification, eliminating traditional preprocessing. With the CWRU dataset, deep learning models, especially CNNs, achieved over 95% accuracy across various loads. The approach simplifies diagnostics and shows strong potential for real-world machinery health monitoring. Yubin Yoo et al. [9] introduces a lightweight CNN model for bearing defect diagnosis, using downsampled vibration signals and spectrograms to reduce data dimensionality. Tested on the CWRU dataset, the model achieves high accuracy with low-dimensional input, outperforming state-of-the-art methods while cutting computation. This approach could extend to other high-dimensional time series analysis fields. Banumalar et al. [10] uses ML techniques on a CWRU vibration dataset to diagnose rolling element bearings. After preprocessing to handle noise and normalize data, feature extraction and model training were conducted. Results show that the method effectively maintains bearing health and supports safe induction motor operation.

In order to discover improved accuracy and shorter processing times, researchers employed a variety of models, according to the literature. Current bearing fault analysis methods face challenges like data imbalance, high computational cost, limited generalizability, overfitting, complex feature engineering, and lack of interpretability, limiting their effectiveness in real-world applications. This work advances recent studies by achieving high accuracy (98%) with a simple, efficient feedforward ANN, outperforming complex models in computational efficiency *(Training time less than 1 second per epoch)*. It introduces robust preprocessing with statistical feature extraction, standardization, and SMOTE to address data imbalance. Tested on the CWRU dataset, the model is adaptable to varied conditions, ensuring practical applicability for real-world predictive maintenance. Literature demonstrates the rarity of certain combinations. Moreover, This study bears considerable importance for Bangladesh's national development. As the country sees tremendous industrial growth, the dependability and productivity of machines become paramount.

2. Theoretical background

An artificial neural network (ANN) is a computational model inspired by the human brain, consisting of interconnected layers of neurons that learn patterns by adjusting weights during training, typically through optimization methods like gradient descent. With its input, hidden, and output layers, an ANN excels at recognizing complex, non-linear relationships, making it valuable in areas like image recognition, speech processing, and predictive modeling [11].

For the CWRU bearing dataset, statistical features were extracted from raw data to simplify and highlight essential patterns. The ANN, trained on these features, effectively classified the data by learning complex relationships, resulting in robust performance

3. Methodology

Dataset Review

The basic unit of a neural network of machine learning (ML) is data. According to recent study, researchers used publicly available datasets that are often accessible online. To verify the efficacy of the methods, they evaluate various current models, algorithms, and technique. The CWRU dataset [12], created by Case Western Reserve University, is a key resource for bearing fault detection research.

Using a 2 HP induction motor with a torque transducer and dynamometer, data was gathered at 12,000 and 48,000 samples per second, with faults added to SKF bearings through electric discharge. The dataset includes 161 records covering both fault-free and faulted data under different motor loads and speeds, with details on defect type, location, size, and load. The data, initially in MATLAB format and later converted to CSV, is highly valuable for building and testing predictive maintenance models. Other similar datasets, like MFPT and IMS, also support fault detection studies, but rely more on accelerated testing. The CWRU dataset stands out for its detailed, accessible data, making it essential for machine learning applications in predictive maintenance and motor health monitoring. The experimental setup of CWRU is shown in **Fig 1.** The spectrum of data at different working condition are shown in **Fig 2.**

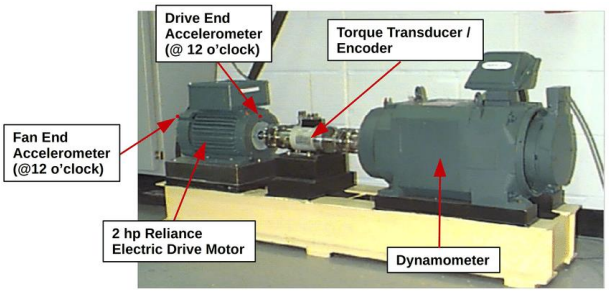


Fig 1. CWRU bearing setup [13]

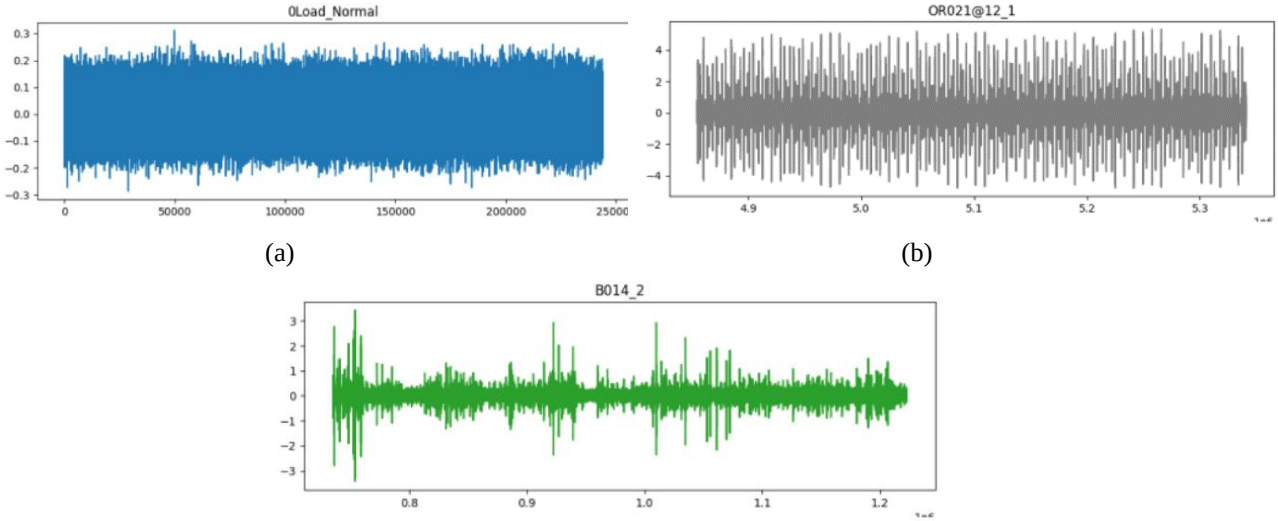


Fig 2. (a) Spectrum of Normal condition at 0 HP load (b) Spectrum of Outer race fault at 1 HP load (c) Spectrum of Ball fault at 2 HP load [1]

In this experiment, 2 HP motor tested with a 1 HP load and a rotation speed of 1772 RPM [12]. Fault Types were single-point defects introduced by EDM (Electric Discharge Machining) on the ball, inner race, or outer race with diameters of 0.007, 0.014, and 0.021 inches [12]. The accelerometers sampled data at 48 kHz to capture fine details of motor vibrations under fault conditions [12]. Measurements were taken from the Drive End (DE), Fan End (FE), and Base (BA) of the motor [12].

This dataset includes nine statistical features calculated over time segments of 2048 data points (equivalent to 0.04 seconds at the 48 kHz sampling rate). These features capture key characteristics of motor vibrations and are essential for identifying faults. The nine features are: (1) Maximum: The peak value in each segment, indicating the highest point of vibration. (2) Minimum: The lowest value in each segment, providing contrast to the maximum. (3) Mean: Average value, giving an overall signal level for each segment. (4) Standard Deviation: Measures the spread or variability in the segment. (5) Root Mean Square (RMS): Shows the energy of the signal, useful for assessing signal strength. (6) Skewness: Indicates any asymmetry in the signal distribution. (7) Kurtosis: Highlights any peakedness or flatness in the distribution. (8) Crest Factor: Ratio of the peak to RMS, helpful for detecting abnormal peaks. (9) Form Factor: Ratio of RMS to the mean, which helps in identifying fault patterns.

These features are well-suited for machine learning models as they capture the essential characteristics of each time segment, aiding in fault identification and classification.

Workflow

The CWRU dataset was utilized for statistical feature extraction. The Synthetic Minority Over-sampling Technique (SMOTE) was used to enhance the data. The model is trained using an ANN, and the outcomes are assessed. The workflow is shown in **Fig. 3**.

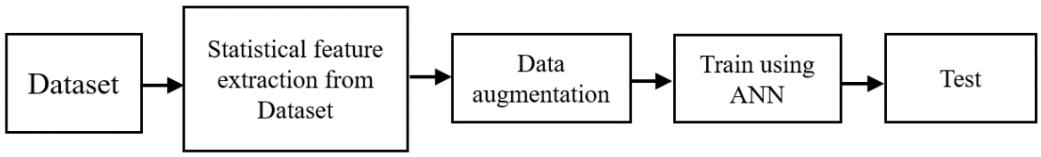


Fig 3. Workflow chart

Data preprocessing

In this study, the dataset initially contained 230 samples per class, with 10 classes, totaling 2,300 samples. The features consisted of 9 statistical values calculated from motor performance data, including maximum, minimum, mean, standard deviation, RMS, skewness, kurtosis, crest factor, and form factor. To address the limited amount of data in our dataset, augmentation was performed applying SMOTE. This technique generated synthetic samples for the minority classes by interpolating between existing samples. The augmentation increased the dataset size to 10,000 samples (1,000 per class), ensuring that each class had an equal number of instances, which is critical for improving model training and performance. We standardized the features using Standard Scaler. This method centers the data by removing the mean and scaling it to unit variance. This step was essential to ensure that all features contributed equally to the model and to avoid bias toward variables with larger ranges

$$x_{scaled} = \frac{x - \mu}{\sigma} \tag{1}$$

Where, μ is the mean of the feature and σ is the standard deviation of the feature. The labels were one-hot encoded to transform the categorical labels (10 classes) into a binary matrix. Each label was represented as a vector of length 10, where only one element was "1" (indicating the class), and all other elements were "0". This encoding is commonly used for multi-class classification problems to ensure the model correctly learns the class associations. By this way label encoding was completed.

Proposed Model

The model employed in this work is a feedforward Artificial Neural Network (ANN) designed for a classification task. The network consists of an input layer, three hidden layers, and an output layer. The breakdown of the model architecture is shown in **Fig. 4.** The three layers are:

(1) Input Layer: The network takes 9 features as input. Each feature represents a distinct attribute of the data, which are fed into the network for processing. These features are represented as a vector. In the equation (1) each x_i corresponds to a specific input feature.

$$x = \{x_1, x_2, x_3, \dots\dots\dots, x_9\} \tag{1}$$

(2) Hidden Layers: The model includes three hidden layers, each containing 64 nodes. These layers are fully connected. Each hidden layer uses the ReLU (Rectified Linear Unit) activation function, which helps the model learn complex patterns by introducing non-linearity. The output of each hidden layer is computed as:

$$H_{l,n} = ReLU\left(W_l H_{l-1,n} + b_l\right), l \in \{1,2,3\}, n \in \{1,2,\dots,64\} \tag{3}$$

where W_l represents the weight matrix for layer l, $h_{l-1,n}$ is the input from the previous layer and b_l is the bias term for that layer.

(3) Output Layer: The output layer contains 10 nodes, corresponding to the 10 possible classes. The output of this layer is passed through the softmax activation function, which converts the raw output values into probabilities. The softmax function ensures that the output values are in the range [0, 1] and sum to 1, making them interpretable as class probabilities:

$$y_{pred} = softmax(W_4 H_3 + b_4) \tag{4}$$

where, W_4 is the weight matrix for the output layer, h_3 is the output from the third hidden layer, and b_4 is the bias term for the output layer.

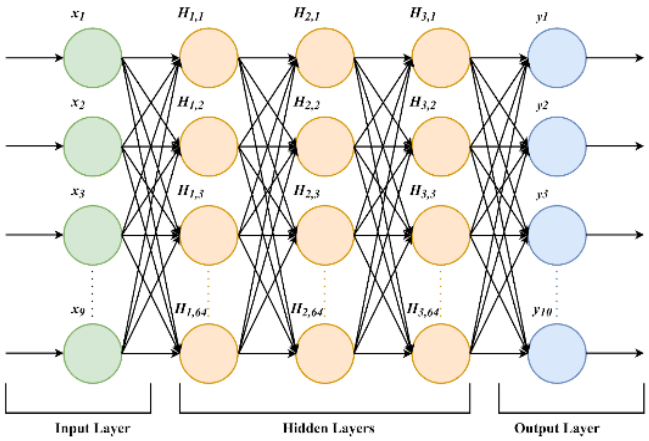
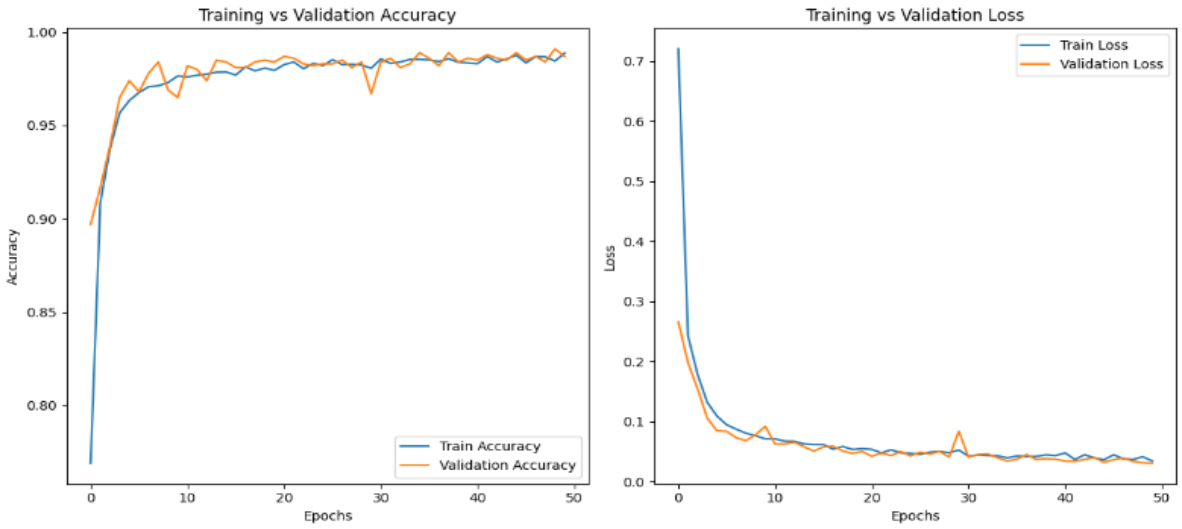


Fig 4. Breakdown of the model architecture

4. Results and Discussion

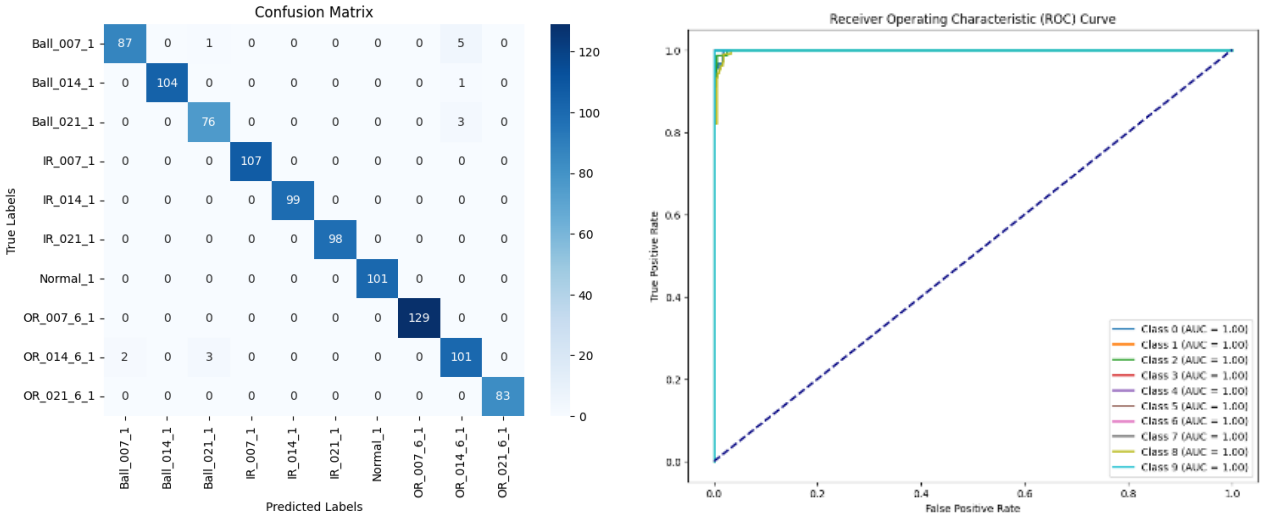
The training and validation accuracy and loss curves (shown in **Fig 5**) display the model's learning behavior across 50 epochs. Training and validation accuracy and loss curves represent the model's performance during training, showing how well it learns (accuracy) and how much error it makes (loss) on both the training data and unseen validation data over time. Both training and validation accuracies approach 1.0, with training and validation losses stabilizing near zero. The parallel trends of these curves suggest that the model has learned effectively from the data without significant overfitting. These results reflect the model's robust learning and generalization ability, showing minimal performance gaps between the training and validation sets. The confusion matrix (shown in **Fig 6(a)**).



(a) (b)

Fig 5. (a) Training vs validation accuracy, (b) Training vs validation loss

provides a detailed view of the model's classification accuracy for each class, highlighting the number of correct and incorrect predictions. A confusion matrix represents the performance of a classification model by summarizing the correct and incorrect predictions across different classes in a tabular format. Most classes show near-perfect classification, with only a few misclassifications. For example, in the "OR_014_6_1" class, some samples were incorrectly classified into other classes, including "Ball_021_1." The ROC curve (Shown in **Fig 6 (b)**) illustrates the model's ability to distinguish between classes, showing a perfect Area Under the Curve (AUC) score of 1.00 for each class. The ROC-AUC curve represents a classification model's ability to distinguish between classes, showing the trade-off between true positive and false positive rates across different thresholds. This high AUC value across all classes suggests that the model effectively identifies true positives without generating false positives. The details of precision recall and F1-Score are shown in **Table 1**. Comparison with recent literature are shown in **Table 2**.



(a) **(b)**

Fig 6. (a) Confusion matrix, (b) Receiver Operating Characteristics (ROC) curve

Table 1. Precision, Recall F1-Score

Class	Precision	Recall	F1-Score
Ball_007_1	0.98	0.94	0.96
Ball_014_1	1.00	0.99	1.00
Ball_021_1	0.95	0.96	0.96
IR_007_1	1.00	1.00	1.00
IR_014_1	1.00	1.00	1.00
IR_021_1	1.00	1.00	1.00
Normal_1	1.00	1.00	1.00
OR_007_6_1	1.00	1.00	1.00
OR_014_6_1	0.92	0.95	0.94
OR_021_6_1	1.00	1.00	1.00
Accuracy			0.98

Table 2. Comparison with recent literature

Works	Year	Model Used	Parameters (10^3)	Training time per epoch	Accuracy (%)
[1]	2024	VGG16	138,357	60-180	99.80
[9]	2023	ResNet50	47,187	120-180	99.97
This work	2024	Custom ANN	9,610	1	98.60

5. Conclusion

This study suggests a feedforward ANN based model to diagnosis and predict bearing fault using CWRU dataset. This model featuring accuracy of 98% with very fast processing time. The bearing vibration data from CWRU at different working condition were extracted into different statistical features. Appling SMOTE the model was trained using ANN. The model demonstrates strong classification performance, achieving perfect AUC scores across all classes and high accuracy in the confusion matrix, with only minor misclassifications. Training and validation accuracy curves closely align, and loss stabilizes near zero, indicating effective learning and minimal overfitting. These results confirm the model's robustness and its ability to generalize well to validation data. Future work could involve testing on more diverse datasets to further assess its generalization capacity. Overall, the model is well-suited for precise class differentiation and demonstrates consistent performance across all metrics.

6. References

[1] D. Saha, M. E. Hoque, and M. E. H. Chowdhury, "Enhancing Bearing Fault Diagnosis Using Transfer Learning and Random Forest Classification: A Comparative Study on Variable Working Conditions," *IEEE Access*, vol. 12, pp. 5986–6000, 2024, doi: 10.1109/ACCESS.2023.3347345.

[2] D. K. Saha, M. E. Hoque, and H. Badihi, "Development of Intelligent Fault Diagnosis Technique of Rotary Machine Element Bearing: A Machine Learning Approach," *Sensors (Basel).*, vol. 22, no. 3, Feb. 2022, doi: 10.3390/S22031073.

[3] M. 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, 1991, doi: 10.1109/41.107100.

[4] 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, doi: 10.1016/J.APACOUST.2013.07.001.

[5] 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, doi: 10.1016/J.ESWA.2013.01.033.

[6] M. He and D. He, "Deep Learning Based Approach for Bearing Fault Diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May 2017, doi: 10.1109/TIA.2017.2661250.

[7] M. Alonso-González, V. G. Díaz, B. L. Pérez, B. C. P. G-Bustelo, and J. P. Anzola, "Bearing Fault Diagnosis With Envelope Analysis and Machine Learning Approaches Using CWRU Dataset," *IEEE Access*, vol. 11, pp. 57796–57805, 2023, doi: 10.1109/ACCESS.2023.3283466.

[8] K. K. Raj, S. Kumar, R. R. Kumar, and M. Andriollo, "Enhanced Fault Detection in Bearings Using Machine Learning and Raw Accelerometer Data: A Case Study Using the Case Western Reserve University Dataset," *Information*, vol. 15, no. 5. 2024. doi: 10.3390/info15050259.

[9] Y. Yoo, H. Jo, and S.-W. Ban, "Lite and Efficient Deep Learning Model for Bearing Fault Diagnosis Using the CWRU Dataset," *Sensors*, vol. 23, no. 6. 2023. doi: 10.3390/s23063157.

[10] K. Banumalar and P. Balakumar, "Efficient Machine-Learning Model for Bearing Fault Identification Using the CWRU Dataset BT - Control and Information Sciences," I. Thirunavukkarasu and R. Kumar, Eds., Singapore: Springer Nature Singapore, 2024, pp. 489–507.

[11] A. N. Al–Dabbous, P. Kumar, and A. R. Khan, "Prediction of airborne nanoparticles at roadside location using a feed–forward artificial neural network," *Atmos. Pollut. Res.*, vol. 8, no. 3, pp. 446–454, 2017, doi: https://doi.org/10.1016/j.apr.2016.11.004.

[12] "Bearing Data Center." Accessed: Jun. 09, 2024. [Online]. Available: https://engineering.case.edu/bearingdatacenter

[13] A. Shenfield and M. Howarth, "A Novel Deep Learning Model for the Detection and Identification of Rolling Element-Bearing Faults," *Sensors (Basel).*, vol. 20, Sep. 2020, doi: 10.3390/s20185112.