

A Comparative Analysis of Machine Learning Algorithms for Fault Classification in Cylindrical Roller Bearings

Prasanta Kumar Samal

Department of Mechanical Engineering
The National Institute of Engineering [NIE]
Mysuru, Karnataka, India

Department of Mechanical Engineering
JSS Science and Technology University
[JSS&TU], Mysuru, Karnataka
India

Srinidhi R

Department of Mechanical Engineering
JSS Science and Technology University
[JSS&TU], Mysuru, Karnataka
India

Imran M Jamadar

Department of Mechanical Engineering
The National Institute of Engineering [NIE]
Mysuru, Karnataka
India

Vishal Baligar

Department of Mechanical Engineering
The National Institute of Engineering [NIE]
Mysuru, Karnataka
India

With the growing demand for enhanced machinery reliability, the application of artificial intelligence (AI) in fault diagnosis is becoming increasingly important. However, limited research has systematically compared the performance of various machine learning algorithms for bearing fault diagnosis. This paper presents a comparative analysis of five widely-used machine learning algorithms—linear SVM, Gaussian SVM, SVM Kernel, Weighted K-Nearest Neighbors (WKNN), and Artificial Neural Networks (ANN). Faults were induced using wire EDM on test bearings, and vibration data were recorded using a National Instruments data acquisition system with LABVIEW. Results indicate that while weighted KNN demonstrated 100% accuracy in testing, ANN emerged as the most reliable, achieving 100% accuracy in both validation and testing phases. The trained ANN was further employed to predict bearing conditions across ten random datasets, affirming its potential for real-time condition monitoring of industrial machinery.

Keywords: *Vibration data acquisition, Statistical feature extraction, Training validation and testing, Artificial Intelligence, Bearing fault diagnosis.*

1. INTRODUCTION

In the contemporary industrial landscape, machinery has become increasingly complex, often operating under unpredictable and challenging conditions. With this rise in complexity, machinery becomes more prone to breakdowns, and unforeseen failures can lead to unplanned maintenance, creating delays that ripple through interconnected systems. Consequently, monitoring the health of machines is essential to ensure their availability, reduce downtime, and increase customer satisfaction [1,2].

Among various machinery components, rolling element bearings (REB) are critical, playing a key role in mechanical systems while also being a significant cause of failures. Studies have shown that between 45% and 55% of failures in rotating machines are attributed to bearing faults [3]. Thus, detecting faults in bearings has become a major area of research aimed at improving reliability and reducing maintenance costs. Fault detection generally involves three stages: data acquisition, feature extraction, and fault classification [4]. Vibration-based data acquisition is widely used due to its effectiveness in identifying machine faults, as vibration signals carry critical information about the operating condition of the bearings [5]. Feature extraction identifies key patterns in these vibration signals, and machine learning algorithms are then applied for fault classification to determine the fault's nature and severity.

Vibration-based monitoring techniques have been extensively studied, with applications in industries such as aerospace, power generation, and material handling [6]. Vibration arises from various sources, even in new bearings with no defects, due to inherent imperfections like variable compliance [7]. These vibrations can be analyzed in both the time and frequency domains using various advanced signal processing techniques. Methods include analog signal acquisition with accelerometers and statistical analysis, which extract parameters like peak value, root mean square (RMS), kurtosis (KU), and crest factor (CF) to assess the condition of the bearings [8–10]. Additionally, characteristic frequencies such as Ball Pass Frequency of the Outer Race (BPFO), Ball Pass Frequency of the Inner Race (BPFI), and Fundamental Train Frequency (FTF) serve as critical indicators of bearing faults [11].

With advancements in artificial intelligence (AI), machine learning techniques like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have been employed for fault diagnosis. ANNs, particularly using the feed-forward back-propagation (FFBP) algorithm, have been widely used due to their pattern recognition capabilities [12]. In fault diagnosis, SVMs have demonstrated high accuracy when applied to time-domain and frequency-domain data, particularly for cylindrical roller bearings [11]. Moreover, environments like MATLAB are extensively used for employing machine learning algorithms to diagnose faults in REBs [13,14].

Although fault diagnosis in rolling element bearings has been extensively studied, the specific focus on cylindrical roller bearings with roller defects remains relatively underexplored. This gap exists because analyzing vibration signals from cylindrical roller bearings is

Received: December 2024, Accepted: October 2025

Correspondence to: Prasanta Kumar Samal
Department of Mechanical Engineering, The National
Institute of Engineering, Mysuru, Karnataka, India.

Email: prasantaku.samal@gmail.com

doi: 10.5937/fme2504681K

© Faculty of Mechanical Engineering, Belgrade. Allrights reserved

FME Transactions (2025) 53, 681-692 681

often complicated by background noise, which makes the detection of roller-specific faults particularly challenging [15–17]. Furthermore, while artificial intelligence (AI) techniques have been employed in fault diagnosis for rolling element bearings, to the best of the authors' knowledge, there is currently no literature that provides a comparative study of various machine learning algorithms aimed at conclusively demonstrating their performance in condition monitoring.

Therefore, an attempt has been made to fill these gaps by presenting a comparative analysis of five top performing machine learning algorithms for fault classification in cylindrical roller bearings. The primary contributions of this work are:

1. **Five top performing algorithms:** Five widely used machine learning algorithms—linear SVM, Gaussian SVM, SVM Kernel, Weighted K-Nearest Neighbors (WKNN), and ANN—were evaluated in MATLAB for their performance in terms of validation and testing accuracy.
2. **Experimentation and data acquisition:** The experiments were conducted using a machinery fault simulator on SKF cylindrical roller bearings (N204 ECP), which included both healthy and faulty conditions in the outer race and rollers at varying speeds. Faults were introduced using wire Electrical Discharge Machining (EDM), and vibration signals were recorded using a National Instruments data acquisition system with LABVIEW.
3. **Feature extraction and performance evaluation:** Statistical features such as kurtosis and crest factor were extracted from the vibration signals using MATLAB. These features served as inputs for training, validating, and testing the machine learning algorithms.
4. **Prediction by the best algorithm:** Among the tested algorithms, the ANN with a feed-forward back-propagation algorithm performed the best, achieving a validation and testing accuracy of 100%. This algorithm was further validated using ten random datasets to ensure robustness and consistency in fault prediction.

The remaining part of the paper consists of the following sections: In Section 2, the methodology, including experimentation and data acquisition, feature extraction, and training of the machine learning algorithms, is given. Section 3 elaborates on the obtained results and the conclusions in Section 4.

2. METHODOLOGY

Figure 1 depicts the methodology flowchart employed in this investigation. The experiments have been conducted using a machinery fault simulator, to acquire bearing vibrations. The statistical features have been extracted from the raw data using MATLAB to form a comprehensive dataset. This dataset has been used as input for the machine learning algorithms for their training, testing, and validation.

2.1 Vibration data acquisition from test bearings

The experiments have been conducted using a machinery fault simulator, as illustrated in Figure 2, with

different bearing conditions; both healthy states and faults in the outer race and rollers. The experimental setup comprises a 0.25 HP electric motor, a flexible coupling to compensate for misalignment between the motor and shaft, and two support bearings within pedestals. The good bearing is positioned at the near end, while the test bearing is mounted at the far end of the motor. The motor shaft's speed is controlled by a variable frequency drive (VFD). The flow chart of experimentation and data acquisition is shown in Figure 3.

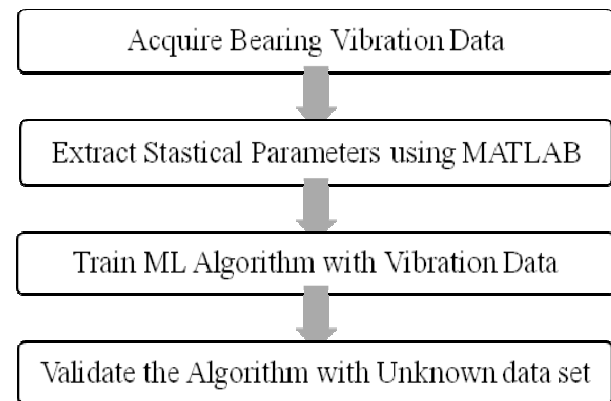


Figure 1. Flow Chart of the Methodology

An accelerometer is mounted to the pedestal to capture the vibration of the test bearing, as shown in Figure 2. This accelerometer has been connected to the NI9234, a specialized module made especially for acquiring vibration and sound, has been mounted on to the NI cDAQ 9178 chassis. LabVIEW facilitates data collecting, and a USB cable connects this chassis to the PC. A LabVIEW program, outlined in Figure 4, has been developed. Initially, the vibration signal in the time domain is acquired, followed by conversion to the frequency domain using the 'spectral measurement' tool in LabVIEW. Subsequently, both time domain and frequency domain data are plotted and saved using the 'write to measurement' tool for further analysis. The saved time domain data will be utilized to extract statistical features using MATLAB.

The specifications of the test bearing (SKF N204: cylindrical roller bearing) are presented in Table 1. Different faults have been induced using wire EDM on the test bearings, as illustrated in Figure 5. Experiments have been conducted for each bearing under four different speeds, as detailed in Table 2. Each test condition was repeated five times, and the vibration data were recorded for further analysis. Consequently, a total of 60 experiments (3 bearing conditions $\times$ 4 speeds $\times$ 5 repetitions) have been carried out.

Table 1. Specification of the Test Bearings (N204)

Type	Cylindrical Roller Bearing
Inner diameter (ID)	20 mm
Outer diameter (OD)	47 mm
Race width (B)	14 mm
Roller diameter (d)	6 mm
Pitch diameter (D)	35.99 mm
No. rollers (n)	10
Contact angle(ϕ)	0°
Material	Chrome Steel

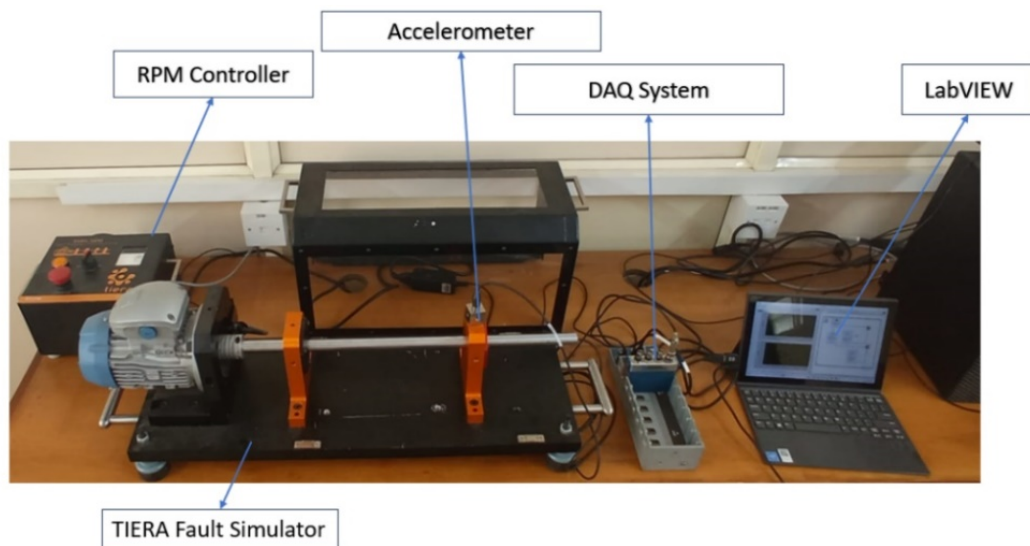


Figure 2. Experimental setup

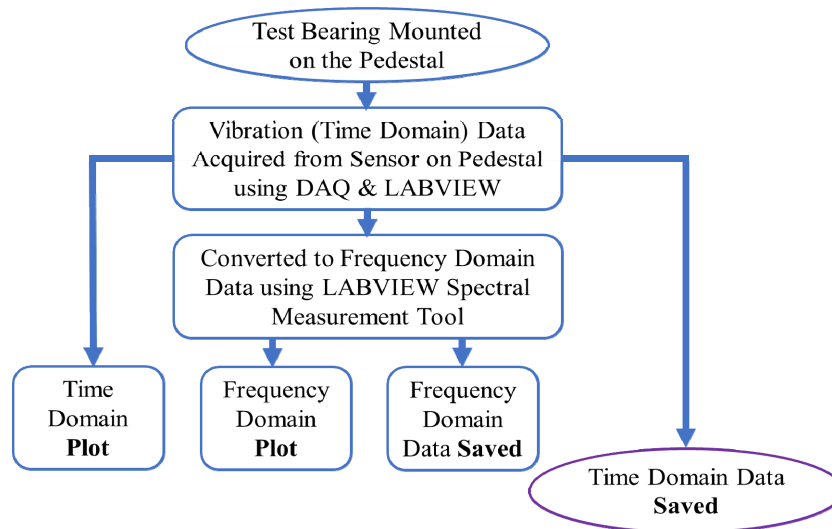


Figure 3. Flow Chart of Experimentation and Data Acquisition

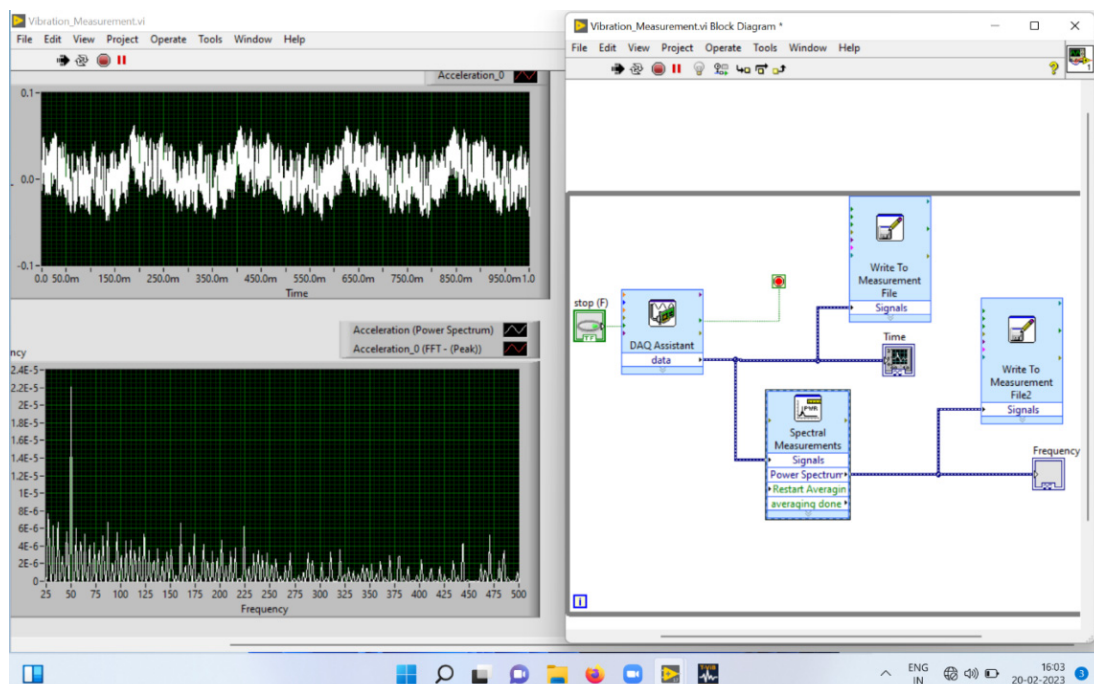


Figure 4. Data Acquisition using LabVIEW

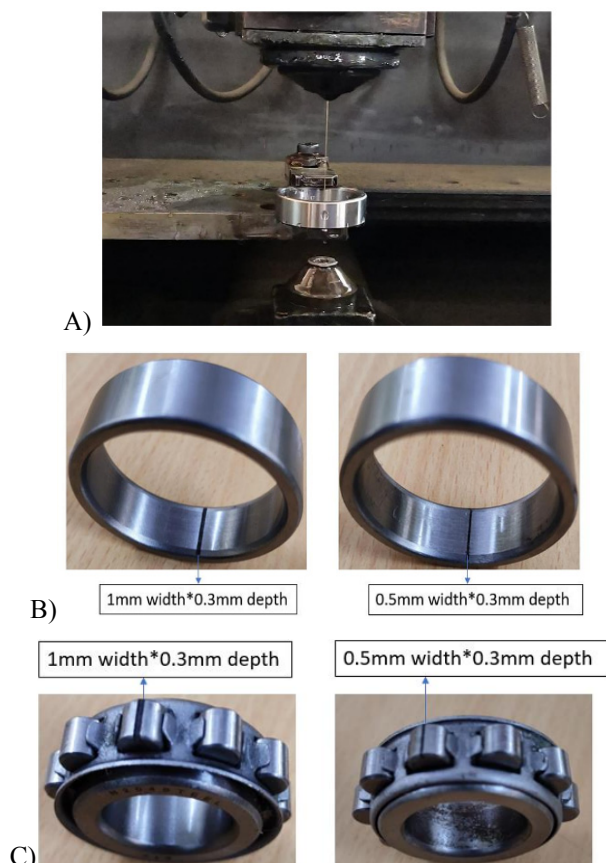


Figure 5. Faults on the Test Bearings

Table 2. Test Conditions

Test Bearing Condition			Speed (RPM)
Healthy Bearing (HB)	Fault on Outer Race (ORD)	Fault on Roller (RED)	100
			200
			300
			400

2.2 Statistical feature extraction using MATLAB

The time-domain signal can be utilized in identifying the faults through statistical features like RMS, Kurtosis, and Crest Factor [18]. Notably, indicators like KU and CF exhibit heightened sensitivity in the diagnosis of faults in rolling element bearings [19]. RMS value,

Kurtosis and Crest Factor can be computed using (1), (2) and (4), with standard deviation using (3) [20,21].

$$\text{RMS} = \sqrt{\frac{y^2}{2}} \approx 0.707y \quad (1)$$

$$\text{KU} = \frac{\sum_{i=1}^N (y_i - \bar{y})^4}{N * \sigma^4} \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N}} \quad (3)$$

$$\text{CF} = \frac{\text{Peak Value}}{\text{RMS Value}} \quad (4)$$

Statistical features, namely RMS, kurtosis, and crest factor have been extracted using the Diagnostic Feature Designer tool in MATLAB from the vibration data to form a robust data set for training and testing the machine learning algorithms.

2.3 Training, validation, and testing of machine learning algorithms

Machine Learning (ML) has gained immense popularity for its ability to empower computers with the capacity to learn patterns and make predictions or decisions without explicit programming. This capability is realized through the application of algorithms and statistical models. MATLAB, a high-level programming language and environment developed by MathWorks, finds widespread usage in scientific and engineering applications, including machine learning [22–25]. Figure 6 shows the flow chart of the training, validation, and testing of five widely used machine learning algorithms for bearing fault classification.

In this analysis, five prominent machine learning algorithms within the MATLAB environment—Linear SVM, Gaussian SVM, SVM Kernel, Weighted KNN, and ANN—have been explored. These algorithms are trained using a comprehensive dataset prepared as detailed in the preceding section. The primary objective is to accurately differentiate between healthy bearings and those with outer race or roller defects.

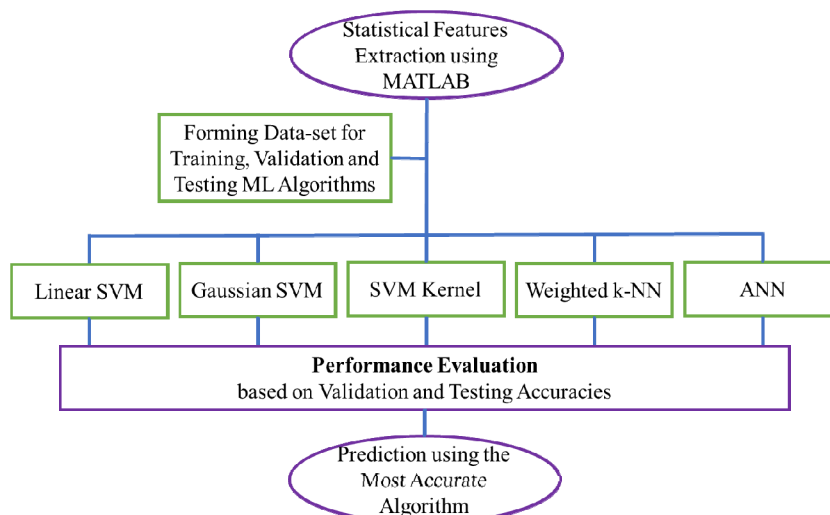


Figure 6. Flow Chart of Training, Validation and Testing of Machine Learning Algorithms

3. RESULTS AND DISCUSSION

3.1 Vibration signals from test bearings

A peak in the frequency spectra indicates the existence of a problem in a bearing element. However, the size, shape, position, and kind of the fault all affect how much this vibration peak vibrates. The roller pass frequency for the outer and inner races (RPFO/RPFI), the roller spin frequency (RSF), and the fundamental train frequency (FTF) are the usual labels for the bearing characteristic frequencies for the faults on the outer race, inner race, and cage, respectively.

Equations (5–10) [26] illustrate how the shape of the bearing and the shaft's rotational speed affect these theoretical bearing characteristic frequencies. Using the (5–10), the theoretical bearing characteristic frequencies for the test bearing with the geometric features from Table 1 for a shaft speed (N) of 400 rpm have been calculated and are shown in Table 3.

$$RPFO = \frac{n}{2} \frac{N}{60} \left(1 - \frac{d}{D} \cos \varphi \right) \quad (5)$$

$$RPFI = \frac{n}{2} \frac{N}{60} \left(1 + \frac{d}{D} \cos \varphi \right) \quad (6)$$

$$RSF = \frac{D}{2d} \frac{N}{60} \left(1 - \frac{d^2}{D^2} \cos^2 \varphi \right) \quad (7)$$

$$FTF = \frac{1}{2} \frac{N}{60} \left(1 - \frac{d}{D} \cos \varphi \right) \quad (8)$$

$$REDF = 2 * RSF = \frac{D}{d} \frac{N}{60} \left(1 - \frac{d^2}{D^2} \cos^2 \varphi \right) \quad (9)$$

$$VCF = n * FTF = \frac{n}{2} \frac{N}{60} \left(1 - \frac{d}{D} \cos \varphi \right) \quad (10)$$

Vibration signals were acquired from bearings using the machine fault simulator and LabVIEW for three bearing conditions: healthy, fault on outer race, and fault on roller. The simulator operated for 10 seconds at four different speeds—100 RPM, 200 RPM, 300 RPM, and 400 RPM. For each type of bearing, 10 data sets

have been recorded to ensure a robust dataset. Figures 7 – 12 show the time domain and frequency domain plots for the three different bearings.

Table 3. Characteristic Frequencies of Test Bearing

Characteristic Frequencies (Hz)	
Variable Competence Frequency (VCF)	27.78
Roller Pass Frequency for the Outer Race (RPFO)	27.78
Roller Pass Frequency for the Inner Race (RPFI)	38.89
Roller Spin Frequency (RSF)	19.44
Fundamental Train Frequency (FTF)	2.778
Rolling Element Defect Frequency (REDF)	38.88

Different peaks may be seen in Figure 8 at roughly 223 Hz, 247 Hz, and 277 Hz. These frequencies are linked to the variable competence frequency (VCF) harmonics. The VCF in theory is 27.78 Hz, which suggests a sound bearing. Conversely, the observation in Figure 10 shows peaks at 112 Hz, 246 Hz, and 271 Hz, aligning with the harmonics of the RPFO. The anticipated theoretical RPFO is 27.28 Hz, indicating the presence of a fault in the outer race. Furthermore, Figure 12 exhibits peaks nearly at 114 Hz and 272 Hz, correspond to the harmonics REDF. The theoretical REDF is 38.88 Hz, pointing towards a fault in the roller. These observations provide valuable insights into the specific nature of the faults present in the bearing components.

3.2 Statistical features from the acquired vibration signals

The statistical features, namely RMS, crest factor, and kurtosis, have been extracted from the saved vibration data using MATLAB. A sample data of the statistical features extracted from the vibration signal is presented in Table 4.

In Table 4, it can be noted from column 4 that kurtosis values are approximately 3 for healthy bearings, while higher values indicate faults in the bearings. Furthermore, it is observed that the kurtosis value remains consistent across different speeds, aligning with the findings reported by previous literatures [26–27].

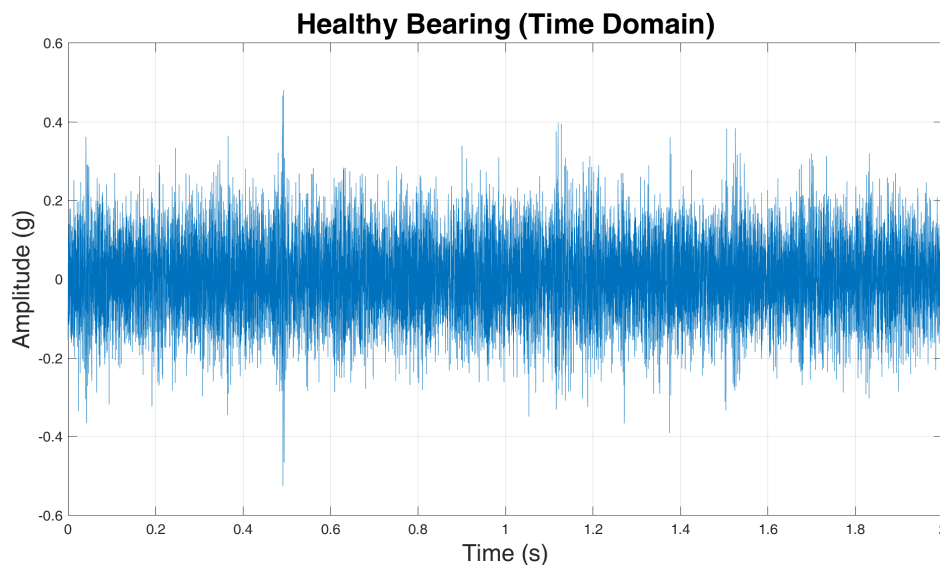


Figure 7. Time Domain Plot for the Healthy Bearing

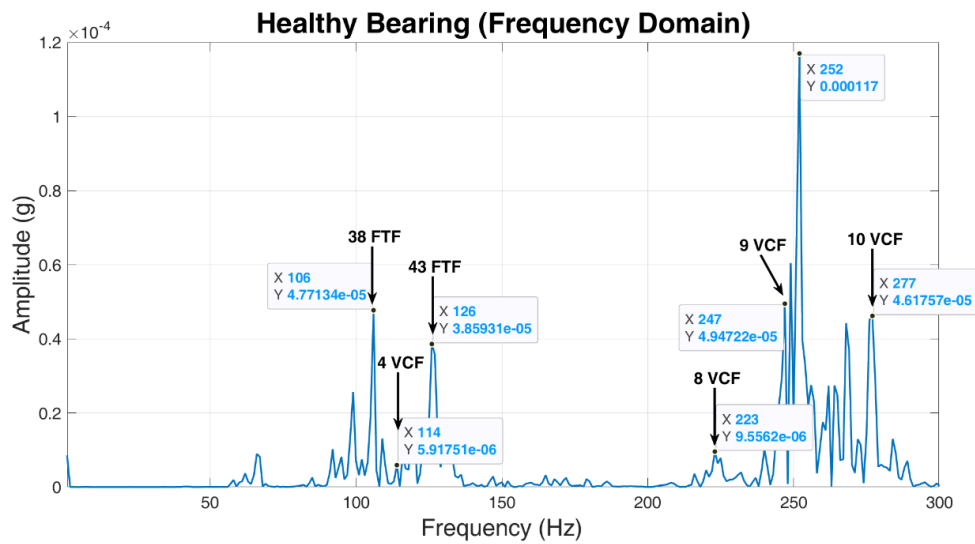


Figure 8. Frequency Domain Plot for the Healthy Bearing

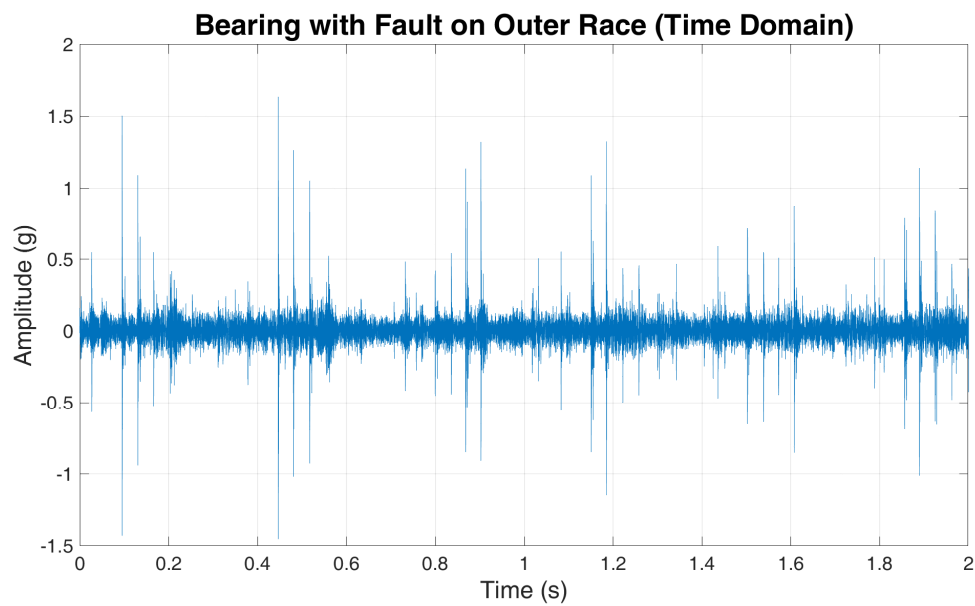


Figure 9. Time Domain Plot for Bearing with Fault on Outer Race

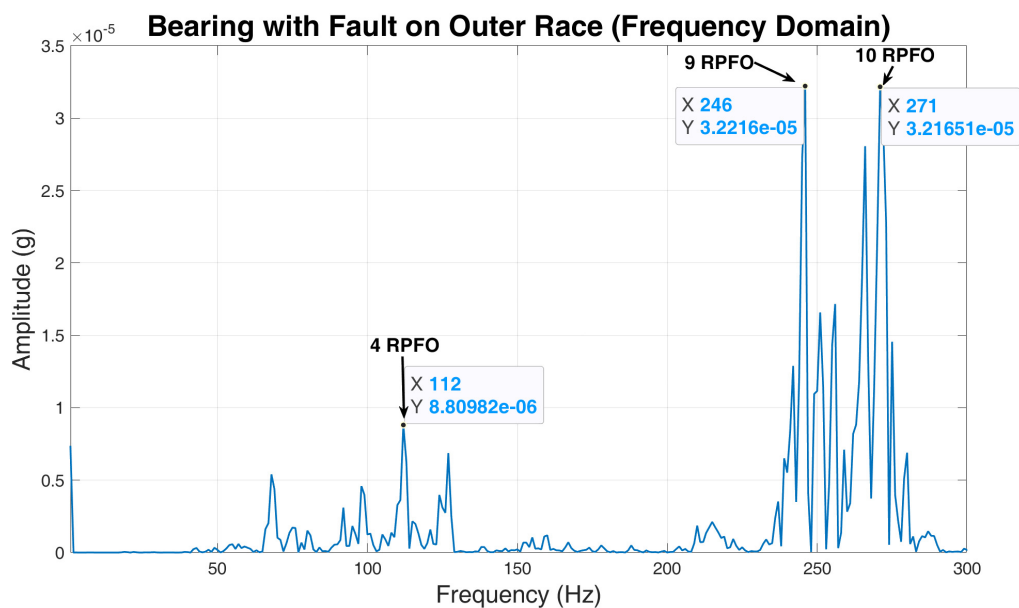


Figure 10. Frequency Domain Plot for Bearing with Fault on Outer Race

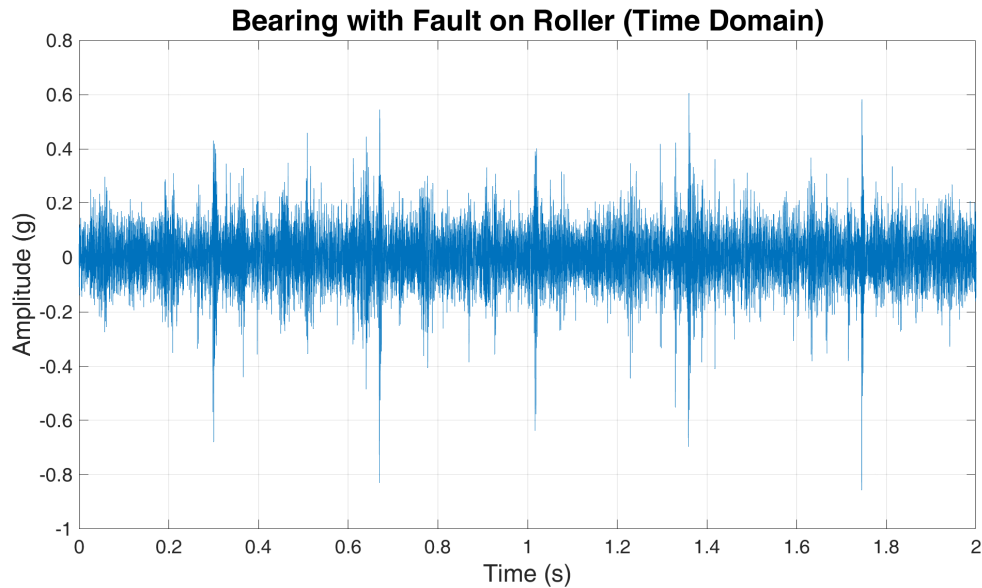


Figure 11. Time Domain Plot for Bearing with Fault on Roller

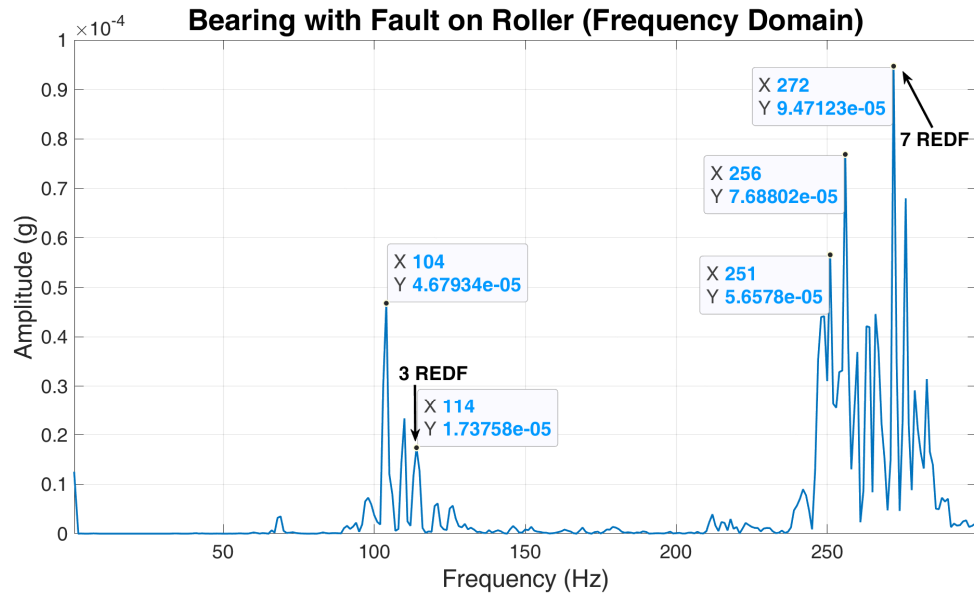


Figure 12. Frequency Domain Plot for Bearing with Fault on Roller

Similarly, examining column 5 of Table 4 reveals that the crest factor values for healthy bearings are approximately 3.5, whereas higher values are indicative of bearings with faults. This observation is consistent with the findings presented in previous literature [28].

Table 4. Sample of Statistical Features

Test Bearing Condition	Speed (RPM)	RMS	Kurtosis	Crest Factor
Healthy (HB)	100	0.006	3.023	3.570
	200	0.009	3.009	3.686
	300	0.010	2.915	3.600
	400	0.013	2.866	3.460
Fault on Outer Race (ORD)	100	0.006	3.626	4.285
	200	0.006	4.085	4.219
	300	0.008	3.847	5.041
	400	0.020	5.508	6.603
Fault on Roller (RED)	100	0.011	7.194	9.330
	200	0.038	5.948	10.362
	300	0.063	5.993	13.811
	400	0.086	6.228	14.207

3.3 Training, validation, and testing of machine learning algorithms

As Heng and Nor [27] have concluded that, employing more advanced parameters on vibration signals did not offer a significant advantage over using kurtosis or crest factor for identifying faults in rolling element bearings.

In this analysis, kurtosis and crest factor have been selected as the input responses for training, validation, and testing of the five machine learning algorithms namely, Linear SVM, Fine Gaussian SVM, SVM Kernel, Weighted KNN, and ANN.

Figure 13 shows the scatter plot (crest factor Vs kurtosis plot) of the input data for three different conditions of bearings namely, healthy bearings (HB) bearings with faults on the outer race (ORD) and bearings with faults on the roller (RED). The validation and testing performance of these algorithms are discussed in the following sections, focusing on their ability to distinguish between these conditions.

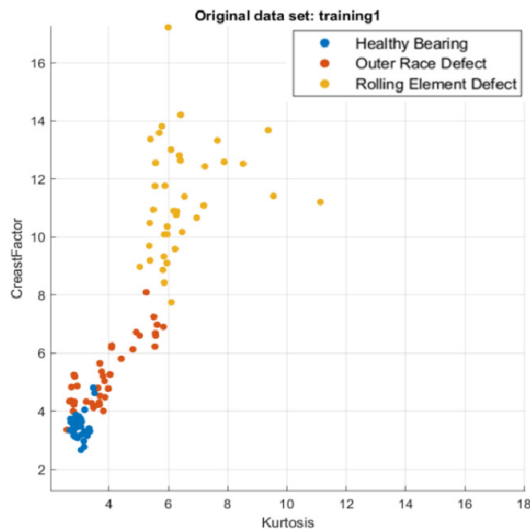


Figure 13. Scatter Plot of the Input Data

3.3.1 Linear SVM

The performance of the linear SVM model for the classification of bearings is analyzed based on scatter plot as shown in Figure 13 and validation and testing confusion matrices as presented in Table 5.

In the confusion matrices the rows represent the true class, and columns depict the predicted class. Correct classifications are indicated by blue-background squares on the diagonal, while misclassifications are represented by light pink background squares off the diagonal. The corresponding percentages of correct classifications and misclassifications are calculated and detailed in Table 5.

Table 5. Confusion Matrices for Linear SVM Model

Validation Confusion Matrix					Testing Confusion Matrix				
True Class	HB	92.5	7.5	0	True Class	HB	92.5	7.5	0
	ORD	7.5	92.5	0		ORD	5	95	0
	RED	0	2.6	97.4		RED	0	0	100
		HB	ORD	RED			HB	ORD	RED
Predicted Class					Predicted Class				

Validation Confusion Matrix Analysis: In the validation matrix, the first-row, first-column square indicates that 92.5% samples were correctly classified as healthy bearings (HB). However, the first-row, second-column square reveals that 7.5% samples were misclassified as bearings with an outer race fault (ORD) when they were, in fact, healthy. Similar observations apply to other classes, highlighting certain misclassifications.

Testing Confusion Matrix Analysis: The testing matrix exhibits improved predictions compared to the validation matrix. Notably, all 100% samples of bearings with roller faults (RED) were correctly classified. However, challenges persist, such as misclassifying healthy bearings as those with outer race faults.

Conclusion: The correlation between confusion matrices and the scatter plot, as evidenced by Figure 13, highlights the interpretability of the model's classification behavior. The distinct placement of samples with roller faults in the scatter plot, juxtaposed with the overlap in other classes, aligns seamlessly with the confusion matrix results. This visual confirmation provides a

clear understanding of the model's occasional challenges, particularly in distinguishing between healthy bearings and those with outer race faults. Furthermore, the comparison between the testing and validation confusion matrices, presented in Table 5, firmly establishes the model's improved predictive capabilities, a testament to its adaptability. This enhancement can be attributed to the meticulous fine-tuning and hyper parameter adjustments carried out during the validation process, ensuring optimal performance in real-world scenarios.

3.3.2 Fine Gaussian SVM

The performance of the Fine Gaussian SVM model for the classification of bearings is analyzed based on scatter plot (Figure 13) and validation and testing confusion matrices as shown in Table 6.

Table 6. Confusion Matrices for Fine Gaussian SVM Model

Validation Confusion Matrix					Testing Confusion Matrix				
True Class	HB	92.5	7.5	0	True Class	HB	97.5	2.5	0
	ORD	5	92.5	2.5		ORD	2.5	97.5	0
	RED	0	0	100		RED	0	0	100
		HB	ORD	RED			HB	ORD	RED
Predicted Class					Predicted Class				

Validation Confusion Matrix Analysis: In the validation matrix, the first-row, first-column square indicates 92.5% samples correctly classified as healthy bearings (HB). However, the matrix also reveals misclassifications, such as 3 samples (7.5%) misclassified as bearings with faults on the outer race (ORD). Similar observations apply to other classes, emphasizing specific misclassification instances.

Testing Confusion Matrix Analysis: The testing matrix demonstrates improved predictions, with 39 samples (97.5%) correctly classified as healthy bearings (HB). Nonetheless, there are still instances of misclassification, as seen in the matrix. Notably, the model could correctly classify all 39 samples (100%) of bearings with roller faults (RED).

Conclusion: The relationship between the scatter plot (Figure 13) and confusion matrices (table 6) is evident. The clear arrangement of samples with roller faults, juxtaposed with the observed overlap in other classes, harmonizes well with the predictions reflected in the confusion matrices. This alignment contributes to a holistic comprehension of the model's occasional challenges, particularly in its ability to distinguish between healthy bearings and those with outer race faults.

Further, the comparison between testing and validation confusion matrices in Table 6 establishes the model's improved predictive capabilities. This improvement is due to careful fine-tuning and hyperparameter adjustments during the validation process, ensuring optimal performance in real-world situation.

3.3.3 SVM Kernel

The performance of the SVM Kernel model for the classification of bearings is analyzed based on scatter plot as shown in Figure 13 and validation and testing confusion matrices as shown in Table 7.

Validation Confusion Matrix Analysis: In the validation confusion matrix, the first-row, first-column square indicates that 38 samples (95%) have been correctly classified as healthy bearings (HB). Conversely, the first-row, second-column square reveals 2 samples (5%) misclassified as bearings with faults on the outer race (ORD) that were actually healthy. Similar observations apply to other classes, emphasizing specific misclassification instances.

Table 7. Confusion Matrices for SVM Kernel Model

Validation Confusion Matrix					Testing Confusion Matrix				
True Class	HB	95	5	0	True Class	HB	95	5	0
	ORD	2.5	92.5	5		ORD	2.5	95	2.5
	RED	0	2.6	97.4		RED	0	0	100
		HB	ORD	RED			HB	ORD	RED
Predicted Class					Predicted Class				

Testing Confusion Matrix Analysis: The testing confusion matrix demonstrates improved predictions, with 38 samples (95%) each correctly classified for both healthy bearings (HB) and bearings with fault on outer race (ORD). However, there is 100% correct classification without any confusion for bearings with fault on roller (RED). These observations are detailed in the matrix.

Conclusion: The correlation between the scatter plot (Figure 13) and confusion matrices (Table 7) is evident, enhancing the understanding of the model's challenges, especially in distinguishing between healthy bearings and those with outer race faults. Additionally, the comparison between testing and validation confusion matrices in Table 7 highlights the model's enhanced predictive capabilities.

3.3.4 Weighted K-Nearest Neighbors (WKNN)

The performance of the weighted k-NN model for the classification of bearings is analyzed based on scatter plot as shown in Figure 13 and validation and testing confusion matrices as shown in Table 8.

Table 8. Confusion Matrices for weighted k-NN Model

Validation Confusion Matrix					Testing Confusion Matrix				
True Class	HB	95	5	0	True Class	HB	100	0	0
	ORD	7.5	92.5	0		ORD	0	100	0
	RED	0	2.6	97.4		RED	0	0	100
		HB	ORD	RED			HB	ORD	RED
Predicted Class					Predicted Class				

Validation Confusion Matrix Analysis: In the validation confusion matrix, the square in the first row and first column indicates that 38 samples (95%) have been correctly classified as healthy bearings (HB). Conversely, the square in the first row and second column shows that 2 samples (5%) have been misclassified as bearings with faults on the outer race (ORD) that were actually healthy. Similar observations apply to other classes, emphasizing specific misclassification instances.

Testing Confusion Matrix Analysis: Similarly, in the testing confusion matrix, the squares in the diagonal indicate that all the samples (100%) in each class have been correctly classified.

Conclusion: The relationship between the scatter plot (Figure 13) and confusion matrices (Table 8) is evident. The clear arrangement of samples with roller faults, juxtaposed with the observed overlap in other classes, harmonizes well with the predictions reflected in the confusion matrices. This alignment contributes to a holistic comprehension of the model's occasional challenges, particularly in its ability to distinguish between healthy bearings and those with outer race faults.

Moreover, the comparison between the testing and validation confusion matrices, as illustrated in Table 8, establishes the model's remarkable improvement in predictive capabilities.

3.3.5 Artificial Neural Network (ANN)

The ANN model randomly allocated 70% of the samples for training and reserved 15% for both testing and validation each. The performance of the ANN model for the classification of bearings is analyzed based on the confusion matrices as shown Table 9.

Table 9. Confusion Matrices for ANN Model

Validation Confusion Matrix					Testing Confusion Matrix				
Output Class	1: HB	100	0	0	Output Class	1: HB	100	0	0
	2: ORD	0	100	0		2: ORD	0	100	0
	3: RED	0	0	100		3: RED	0	0	100
		1: HB	2: ORD	3: RED			1: HB	2: ORD	3: RED
Target Class					Target Class				

Confusion Matrices Analysis: During the training of the ANN model, it was confused for only one sample: a bearing with a healthy bearing (class 1: HB) is misclassified as a fault on the outer race (class 2: ORD). However, all other samples were correctly classified. During validation and testing, the model was capable of classifying all samples (100%) correctly into their respective classes.

Conclusion: This observation is depicted in Table 9. It demonstrates the robustness of the ANN model in bearing fault classification for predictive maintenance of industrial machinery.

3.4 Comparison of Algorithms' Performance

The performance of the considered models, namely Linear SVM, Fine Gaussian SVM, SVM Kernel, Weighted KNN, and ANN have been analyzed based on validation and testing confusion matrices; their performance accuracies are presented in Table 10 for comparison. The accuracy values are plotted for better visual comparison as shown in Figure 14.

Table 10 and Figure 14 provide a clear observation that the ANN model (represented by bars in green) demonstrates 100% accuracy across all three bearing classes during both the validation and testing phases. In contrast, the remaining four models exhibit lower accuracy during validation, with an improvement observed in the testing phase. Notably, the weighted k-NN model

(depicted by bars in purple) achieves 100% accuracy for all three bearing classes during the testing phase. Consequently, it can be concluded that both the weighted k-NN and ANN models stand out as top-performing algorithms among the five.

Table 10. Comparison of Validation and Testing Confusion Matrices for Five Classification Models

Model	Validation Accuracy (%)			Testing Accuracy (%)		
	HB	ORD	RED	HB	ORD	RED
Linear SVM	92.5	92.5	97.4	92.5	95	100
Fine Gaussian SVM	92.5	92.5	100	97.5	97.5	100
SVM Kernel	95	92.5	97.4	95	95	100
WK-NN	95	92.5	97.4	100	100	100
ANN	100	100	100	100	100	100

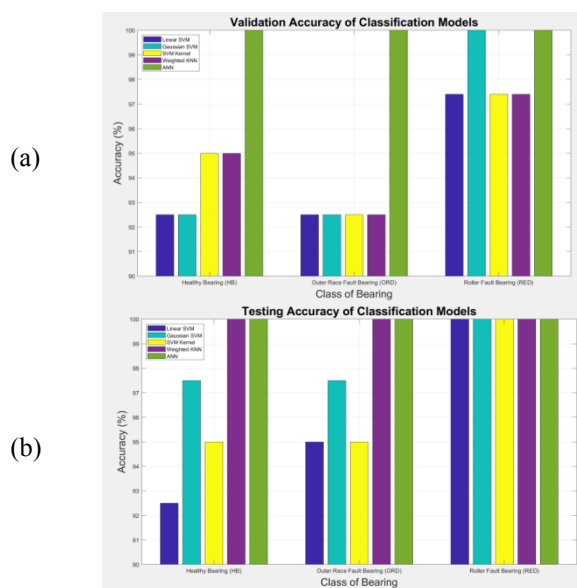


Figure 14. Accuracy of Five Classification Models (a) Validation Accuracy, (b) Testing Accuracy

It is crucial to carefully consider the validation and testing confusion matrices for all five models. Notably, except for the ANN model, the other four models utilize all samples for both validation and testing. In contrast, the ANN model allocates distinct samples—15% each—for validation and testing after utilizing 70% of all samples for training. Even when presented with unseen samples, the ANN model maintains a remarkable 100% accuracy for both validation and testing. Consequently, among the five models investigated in this research, the ANN model emerges as the most reliable and robust choice.

3.5 Prediction of Bearing Condition using ANN

Table 11 shows the details of bearing condition prediction by the trained ANN model.

An additional code has been developed in MATLAB to validate the ANN model. The input consisted of a set of kurtosis and crest factor values, while the output determined the condition of the bearing, specifically, healthy bearing (HB), bearing with a fault on the outer

race (ORD), or RED. The model has been validated using ten random sets of data, revealing accurate classification of bearing conditions. The details of this validation are presented in Table 11.

Table 11. Prediction of Bearing Condition using ANN

Sl. No.	Original Bearing Condition	Input Data		ANN Predicted Bearing Condition
		Kurtosis Value	Crest Factor Value	
1	HB	3.1759	4.0533	HB
2	HB	2.9996	3.7929	HB
3	HB	2.8483	3.5847	HB
4	ORD	3.8644	4.4813	ORD
5	ORD	3.4739	4.585	ORD
6	ORD	3.4738	4.1208	ORD
7	RED	5.3948	10.9439	RED
8	RED	5.5664	10.7667	RED
9	RED	14.1918	21.6435	RED
10	RED	9.985	12.365	RED

The results indicate that when the model has been provided with new and unfamiliar data for prediction, it consistently made correct predictions for all instances in the random datasets, achieving a remarkable 100% accuracy. This high accuracy rate underscores the exceptional performance of the ANN model in classifying bearing conditions, emphasizing its efficacy for predictive maintenance of machinery.

4. CONCLUSIONS

In this research an attempt has been made to conduct a comparative evaluation of five commonly used machine learning algorithms in the MATLAB environment: linear SVM, Gaussian SVM, SVM Kernel, weighted K-Nearest Neighbours, and Artificial Neural Networks. Following conclusions can be drawn from the research:

- The ANN model demonstrates 100% accuracy across all three bearing classes during both the validation and testing phases.
- In contrast, the remaining four models exhibit lower accuracy during validation, with an improvement observed in the testing phase.
- Notably, the weighted k-NN model achieves 100% accuracy for all three bearing classes during the testing phase. Consequently, it can be concluded that both the weighted k-NN and ANN models stand out as top-performing algorithms among the five. However, the ANN model emerges as the most reliable and robust choice.
- Further when the ANN model has been provided with new and unfamiliar data for prediction, it consistently made correct predictions for all instances in the random datasets, achieving a remarkable 100% accuracy.
- The exceptional performance of the ANN model is particularly noteworthy, not only for its accuracy but also for its robustness in handling new, unseen data during the validation and testing phases. This characteristic speaks to the model's potential for practical, real-world applications, especially in the realm of predictive maintenance for industrial machinery.

REFERENCES

- [1] He, Q., Kong, Z. Machine Health Monitoring and Diagnosis: Fundamentals, Tools, and Practices. *Mechanical Systems and Signal Processing*, 145, 106954, 2020.
- [2] Jardine, A. K. S., Lin, D., Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510, 2006.
- [3] Tandon, N., Choudhury, A. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32(8), 469-480, 1999.
- [4] Randall, R. B. *Vibration-based Condition Monitoring: Industrial, Aerospace, and Automotive Applications*. Wiley, 2011.
- [5] El-Thalji, I., Jantunen, E. A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, 60, 252-272, 2015.
- [6] Sassi, S., Badri, B., Thomas, M. A numerical model to predict damaged bearing vibrations. *Journal of Vibration and Control*, 13(11), 1603-1628, 2007.
- [7] Al-Ghamd, A. M., Mba, D. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing*, 20(7), 1537-1571, 2006.
- [8] Smith, W. Accelerometer-based Fault Detection in Rotating Machinery. *IEEE Transactions on Industrial Electronics*, 59(2), 155-165, 2012.
- [9] Saruhan, H., Saridemir, S., Qicek, A., Uygur, I. (2014). Vibration analysis of rolling element bearings defects. *Journal of applied research and technology*, 12(3), 384-395.
- [10] Randall, R. B., Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, 25(2), 485-520, 2014.
- [11] Parmar, U., Pandya, D. H. Experimental investigation of cylindrical bearing fault diagnosis with SVM. *Materials Today: Proceedings*, 44, 1286-1290, 2021.
- [12] Yuan, L., Lian, D., Kang, X., Chen, Y., Zhai, K. Rolling bearing fault diagnosis based on convolutional neural network and support vector machine. *IEEE Access*, 8, 137395-137406, 2020.
- [13] Kim, S., An, D., Choi, J. H. Diagnostics 101: A tutorial for fault diagnostics of rolling element bearing using envelope analysis in matlab. *Applied Sciences*, 10(20), 7302, 2020.
- [14] Prasanta Kumar Samal, Sunil K, Imran M. Jamadar, Srinidhi R. AI-Enhanced Fault Diagnosis in Rolling Element Bearings: A Comprehensive Vibration Analysis Approach. *FME Transactions* 52.3, 2024.
- [15] Xia, H., Chen, M., Wang, J., Wu, Y. Research on Vibration Signal Analysis of Rolling Bearing Based on Transfer Function and Its Application. *IEEE Access*, 10, 1234-1245, 2022.
- [16] Amarnath, M., Sharma, R., Kumar, A. Fault Diagnosis of Rolling Element Bearings: A Review of Techniques and Future Directions. *Journal of Mechanical Science and Technology*, 34(4), 1635-1647, 2020.
- [17] Kim, S., Lee, H., Park, J. Real-time Condition Monitoring of Cylindrical Roller Bearings Using Machine Learning Algorithms. *Mechanical Systems and Signal Processing*, 202, 111234, 2023.
- [18] Norton, M. P., Karczub, D. G. Fundamentals of noise and vibration analysis for engineers. Cambridge university press, 2003)
- [19] Rai, A., Upadhyay, S. H. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96, 289-306, 2016.
- [20] Lebold, M., McClintic, K., Campbell, R., Byington, C., Maynard, K. Review of vibration analysis methods for gearbox diagnostics and prognostics. In Proceedings of the 54th meeting of the society for machinery failure prevention technology (Vol. 634, p. 16). Virginia Beach, VA, 2000.
- [21] I. Howard, A Review of Rolling Element Bearing Vibration "Detection, Diagnosis and Prognosis", Department of Defence, Melbourne, 1994.
- [22] Liu, R., Yang, B., Zio, E., Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, 108, 33-47, 2018.
- [23] Zheng, J., Yuan, Y., Zhao, H., Deng, W. A novel broad learning model-based semi-supervised image classification method. *IEEE Access*, 8, 116756-116765, 2020.
- [24] Mian, T., Choudhary, A., Fatima, S. An efficient diagnosis approach for bearing faults using sound quality metrics. *Applied Acoustics*, 195, 108839, 2022.
- [25] Chen, H., Miao, F., Chen, Y., Xiong, Y., Chen, T. A hyperspectral image classification method using multi feature vectors and optimized KELM. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 2781-2795, 2021.
- [26] Harris T. A, (2001) Rolling Bearing Analysis, 4th ed. New York, John Wiley & Sons;
- [27] Heng, R. B. W., Nor, M. J. M. Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. *Applied Acoustics*, 53(1-3), 211-226, 1998.
- [28] Kurfess, T. R., Billington, S., Liang, S. Y. Advanced diagnostic and prognostic techniques for rolling element bearings. Condition monitoring and control for intelligent manufacturing, 137-165, 2006.

NOMENCLATURE

B	Race Width
d	Roller Diameter
D	Pitch diameter
N	Shaft Speed
n	No. of Rollers

y	vibration magnitude
y_i	instantaneous magnitude
σ	Standard Deviation
ϕ	Contact angle
$\bar{y}$	mean

Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Networks (ANN)
BP	Back-Propagation
CF	crest factor (CF),
CNN	Convolutional Neural Network
EDM	Electrical Discharge Machining
EI	energy index (EI)
FFNN	feed-forward neural network
FTF	Fundamental train Frequency
HB	Healthy Bearing
ID	Inner Diameter
IF	impulse factor (IF),
KNN	K-Nearest Neighbors
KU	kurtosis (KU)
ML	Machine Learning
OD	Outer Diameter
ORD	Fault on Outer Race
REB	Rolling element bearings (REB)
RED	faults on the roller
REDF	Rolling Element Defect Frequency
RMS	Root Mean Square (RMS),
RPFI	Rolling-Element Pass Frequency Inner Race
RPFO	Rolling-Element Pass Frequency Outer Race
RSF	Rolling-Element Spin Frequency
SVM	support vector machine
VCF	Variable Compliance Frequency
VFD	Variable Frequency Drive

WKNN Weighted K-Nearest Neighbors

КОМПАРАТИВНА АНАЛИЗА АЛГОРИТАМА МАШИНСКОГ УЧЕЊА ЗА КЛАСИФИКАЦИЈУ КВАРОВА У ЦИЛИНДРИЧНИМ ВАЉКАСТИМ ЛЕЖАЈЕВИМА

П.К. Самал, Р. Сриниди, И.М. Цамадар,
Б. Балигар

Са растућом потражњом за побољшаном поузда-
ношћу машина, примена вештачке интелигенције
(ВИ) у дијагностици кварова постаје све важнија.
Међутим, ограничена истраживања су систематски
упоређивала перформансе различитих алгоритама
машинског учења за дијагностику кварова лежајева.
Овај рад представља компаративну анализу пет
широко коришћених алгоритама машинског учења -
линеарног SVM-а, Гаусовог SVM-а, SVM језгра,
пондерисаних K-најближих суседа (WKNN) и
вештачких неуронских мрежа (ANN). Кварови су
индуковани коришћењем жичног EDM-а на тест
лежајевима, а подаци о вибрацијама су снимљени
коришћењем система за аквизицију података
National Instruments са LABVIEW-ом. Резултати
показују да је, док је пондерисани KNN показао
100% тачност у тестирању, ANN се појавио као
најпоузданији, постижући 100% тачност и у фази
валидације и у фази тестирања. Обучена ANN је
даље коришћена за предвиђање стања лежајева у
десет случајних скупова података, потврђујући њен
потенцијал за праћење стања индустријских машина
у реалном времену.