Uniting Machine learning models using voting classifiers in the fault diagnosis of IC engine gearbox

Sairam Sridharan, V. Sugumaran, Naveen Venkatesh S

ABSTRACT

A gearbox is an integral part of the Internal Combustion Engine. Any fault in this part should not go unnoticed. This study dives into the exploration and comparison of various classifiers and machine learning models in addressing the gearbox fault diagnosis. Using vibration based diagnostic techniques, the signals from the gearbox was measured. A healthy gear box and 4 different defect conditions (25% defect, 50% defect, 75% defect, 100% defect) were taken under 3 different load conditions (No load, 9.6 Nm, 13.3 Nm). The vibration signals are captured using the Tri-axial accelerometer and the load conditions were changed with the help of an eddy current dynamometer. Subsequently, the features were extracted from the acquired data through 3 different methodologies: Statistical features, Histogram features and ARMA features. Feature selection process was carried out by J48 decision tree algorithm, the important features were extracted from the decision tree and the irrelevant ones were binned of from the dataset. The selected features were then classified using 8 machine learning classifiers. Auto regressive moving average method (ARMA) and Histogram provided 100% accuracy for cases.so, Voting classifier was applied to the statistical datapoints where there was a scope of improvement and the test accuracy can be improved, the Statistical features for Load 2 (13.3 Nm) showed a 3.33 % improvement in the test accuracy. The best accuracy models were noted so as to provide a proper predictive maintenance model for the gearbox fault diagnosis.

Keywords: Gearbox, Fault Diagnosis, ARMA, Histogram, Statistical, Machine Learning, Voting classifier

1. Introduction

An Internal Combustion (IC) engine is a very integral part of various automobiles, the gear box is a major component of the IC Engine. Its main function is to optimize the transmission or transfer power generated by the IC engine to the wheels, which enables the vehicle to locomote in an efficient manner under various conditions and different speed [1]. The gearbox includes gears, shafts and those gears are toothed wheels which mesh with each other in order to transfer motion and power, the ratio of teeth on the different gears determines the torque and rotation speed of the shafts and hence the wheels. The gearbox has a strong influence on the vehicle's locomotion and also influences the driving experience. Inefficiencies or faults in such gearboxes can lead to compromised functioning, lesser fuel efficiency, diminished acceleration. More than that it can lead to catastrophic failures that makes the vehicle inoperable and possess a greater risk for its occupants, even the pedestrians and other nearby vehicles. So, any fault in the gearbox must be addressed quickly and properly [2], [3]. The fault diagnosis in an IC Engine gearbox can be conducted in many ways such as oil analysis [4], temperature analysis [5], sound analysis [6], acoustic emission analysis [7], power analysis [8] and the vibration analysis [9]that is discussed in this paper as well. For this study, we have used vibrational analysis as it is the most commonly used method for identifying faults in engines, rotary components and motors [10]. The detailed explanation about this vibrational study and the necessary equipment are mentioned in the experimental studies section.

Traditionally, the fault diagnosis process was done by a skilled manual labour. The process is time consuming and the efficiency is not enough. This is one of the main reasons for this study to involve machine learning, so as to reduce time consumption and focus on improving the accuracy of the diagnosis [11]. It not only saves time for the current problem-solving process, but also helps in predictive maintenance of any future problems that might arise.

The main objective is to identify the optimal classifier for the purpose of fault diagnosis of the gearbox. Feature engineering is performed on the dataset to get a crisp dataset with proper and important features. The Fault diagnosis of an Internal Combustion Engine Gearbox involves the idea of identifying the issues a vehicle's gearbox has and analysing them. The gear box is responsible for the power transmission from the engine to the wheels at different torque levels and speeds. It is very important to provide varying power to the wheels while performing certain manoeuvres, like when turning, the rear tires have to rotate slowly with less torque and the parallel tires have to a move bit faster than these rear tires. Any fault in the gearbox can lead to the poor performance of the vehicle, transmission failure and can even cause accidents.

For increasing the dependability and accuracies in this field, several researchers have scrutinized a variety of approaches ranging from traditional models to the recent Deep, CNN and machine learning models in the field of internal combustion engine gearbox fault detection and its possible diagnosis.

Some traditional methods such as k-star algorithms and wavelet transform features, time-frequency analysis [12]. These Wavelet transforms have been used for the fault diagnosis [2]. However, such methods have certain limitations in handling some complex and non-stationary signals. There successful application of the ensemble classification models combined along with the Discrete Wavelet Transform (DWT) analysis for detecting and classifying PQDs in a PV integrated microgrid network [13]. However, DWT assumes that a signal is stationary over the viewed time period, but in real world scenario, in dynamic microgrid networks, the non-stationary signals might affect the feature engineering process and the disturbance detection. Traditional methods rely on manually extracted features from the sensor data. The cited research introduced a novel deep neural network called BiConvLSTM and reveals its superiority over convention deep learning methods such as CNNs, LSTMs [14]. However, analyzing the BiConvLSTM and training those deep architectures can be very challenging due to certain issues like overfitting and vanishing gradients.

The fault diagnosis of the IC engine, especially its gearbox has become one of the most dominant research topics over the past few decades. The olden days saw some components that would help test an IC engine, but could only focus on few or very little parameters and were unable to exactly identify the health of the engine's gearbox, then came advanced sensors and machine learning models that can compute quite a lot in fraction of seconds providing the necessary details (in this case the fault diagnosis of the gear box) with greater accuracy levels. The Machine learning models can process huge amounts of data from sensors and identify some complex patterns associated with the signals from the source

This paper presents a comprehensive methodology that contains several essential elements, it outlines the experimental setup for the test, elucidates the feature extraction process, the feature selection process using tree classifier such as J48 to highlight the important features in the dataset so as to get high accuracy from the model by only choosing the relevant and best features, the other features in the data are deemed insignificant and finally the implementation of classification algorithms on those selected features are carried out [15]. The results from this study are compared with the traditional methods in order to highlight the superiority in the proposed machine learning approach.

This paper explores the application of machine learning classifiers such as Support Vector machine (SVM), k-Nearest Neighbor (KNN), Logistic Model Tree (LMT), Random Forest (RF), Multi-layer Perceptron (MLP), Logistics (LO), j48, Naïve Bayes (NB) and even voting classifiers to create 2,3,4 or 5 ensemble models with different classifier combinations to check if there is any improvement in the Testing accuracy of these models.

By using a proper live experimental set-up, the Fault data were recorded for Load1, Load2 and NO load with five classes of 25% defect, 50% defect, 75% defect, 100% defect, No defect [2]. These defects were formed by breaking the gear teeth according to the requirements. The sensors measure the vibrations produced during each condition, and the data is stored respective to the load condition and defect classes (25,50,70,100, No defect).

The major contributions of this paper can be summarized in the following way:

- Feature extraction in 3 ways: Statistical feature extraction, Histogram Feature extraction and ARMA feature extraction
- Feature selection using J48 decision tree algorithm
- Feature Classification using Machine learning algorithms and voting classifier in order to help increase the accuracy.

Table 1 - Summary of related works

REFERENCE	COMPONENT	TECHNIQUES USED
[16]	Gearbox	Using wavelet transform to convert 1D time-domain signals to 2D time-frequency. Convolutional neural network (CNN) model WT-MLCNN for compound fault diagnosis of gearboxes
[17]	Roller bearing	Statistical Feature Extraction Principal component analysis and Decision tree techniques for feature selection (J48) SVM, PSVM Classifiers

[18]	Piston	Statistical features extraction CWT to analyze the vibrational signals
[19]	Gearbox	Frequency-domain feature extraction Sparse filtering Softmax regression
[20]	Board-level	Artificial neural network (ANN) and Support vector machine (SVMs) used for feature classification Weighted-majority voting for better accuracy
[21]	Planetary Gearbox	Histogram feature extraction J48 Fuzzy classifier ANN Naïve Bayes

Research gaps

- Machine learning techniques provide more accurate and efficient results for Gearbox fault diagnosis than the existing traditional methods.
- Vibrational signals tend to outperform acoustic, oil and temperature signals for the gearbox fault diagnosis. (Because better accuracy was obtained when vibrational signals were used)
- Most of the studies only focus on a single set of features, they either focus on histogram features or ARMA features or statistical alone.

Novelty

- Statistical, Histogram and ARMA features are used in this study
- 8 different classifiers (SVM, KNN, LMT, RF, MLP, LO, J48, NB) are applied on these features for the 3 load conditions.
- Voting classifier is a powerful method for increasing the testing accuracy of any dataset, hence it is used in this paper in order to try and improve the accuracies for certain conditions where it is possible.

The rest of the paper is arranged in the following order: Section 2 describes the experimental studies, Section 3 explains the methodology involved, Section 4 holds the results and discussion and about the classifiers used, Section 5 is conclusion

2. Experimental studies

This section on experimental research tells us how the data was acquired from the laboratory test setup.

2.1 Procedure

The IC engine is kept in running condition (in driver gear 2) till a steady speed is achieved, the data acquisition process can begin once a steady speed is achieved. The Experimental procedures were performed on a four stroke IC engine Gearbox, with an eddy current dynamometer, the dynamometer is used to measure and control rotational speed and torque of the gearbox and operates on the principle of Electromagnetic Induction. The Eddy current dynamometer is useful for performance testing, optimization of gear ratios, durability testing, and exactly helps to create different load conditions. For this study, we have made the use of 3 different load conditions: No Load, Load 1 (9.6 Nm Torque), Load 2 (13.3 Nm Torque). These load conditions are applied to all the fault conditions (25% defect, 50% defect, 75% defect, 100% defect, No defect) of the gears in the gearbox (refer Figure 1).

A Tri-axial accelerometer mounted on the top of the casing of the gearbox is used to acquire the vibrational signals. The schematic representation of the setup is shown Flowchart - 1.

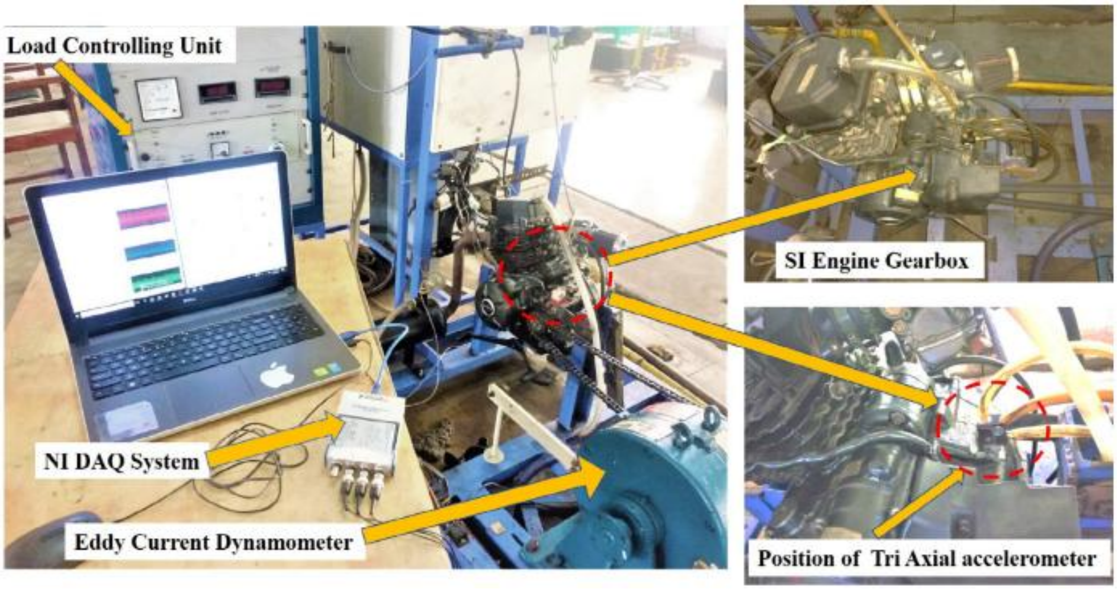


Figure 1 - Experimental setup

The tests were performed on the second driver gear for smoother operation and ensuring a good power transmission for collecting the required data. The gear chosen is initially a healthy one and then is replaced with defective gears for further experimentation. 5 different conditions (Refer Table 3) were chosen, such as 1) Healthy, 2) 25% tooth defect, 3) 50% tooth defect, 4) 75% tooth defect, 5) 100% tooth defect [2]. The gear tooth was broken according to the above given conditions and then the gearbox was made run and the vibrational signals were recorded for all the cases. The Experiment involved 3 different loading conditions, in each load condition, the 4 gears with defects and the healthy one was tested and the vibrational results were recorded. (Refer Figure 2)

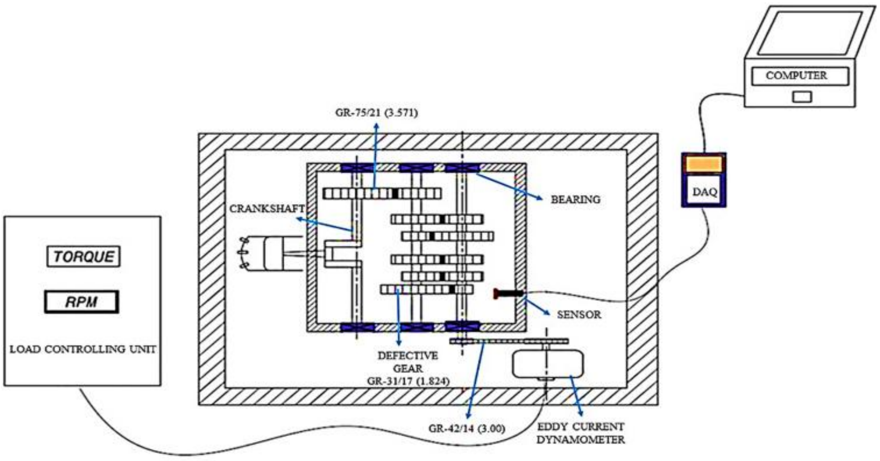


Figure 2 - Sketch of the experimental setup

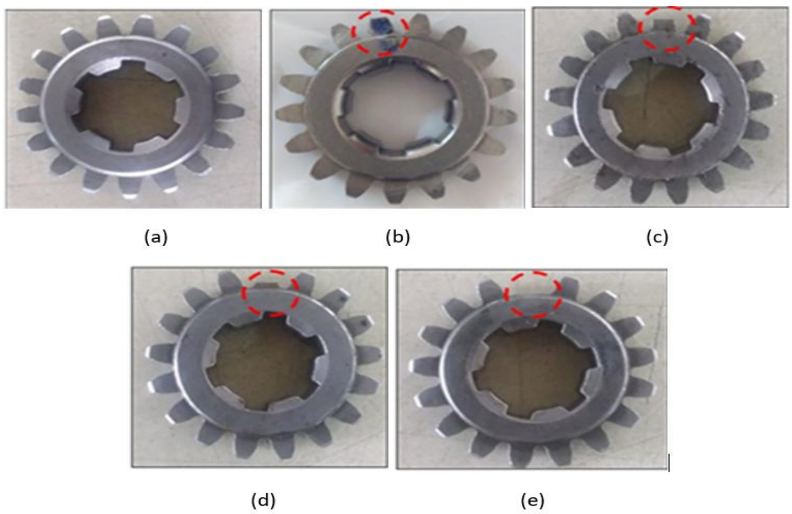


Figure 3 - Gear Tooth Fault Conditions (a) No Defect, (b) 25% Defect, (c) 50% Defect, (d) 75% defect, (e) 100% defect

2.2 Data acquisition

As mentioned above, the Tri axial accelerometer is used to measure the vibration signals from the gear box. For that, a sampling frequency must be set. Sampling frequency is a critical parameter in data acquisition, it refers to the number of samples or data points taken per unit time from a continuous signal. The sampling frequency for this experimental research was taken as 25.6 kHz. This ensures even a minute information within the vibration is preserved and not lost while transmission.

The obtained vibrational signals are analog signals, for performing feature engineering processes and feature classification process, the vibration signals need to be converted into digital signals. A total of 600 samples were recorded, out of which 30 samples were each taken for every fault conditions of the gearbox under all the 3 different load conditions mentioned above.

3. Methodology

The methodology involved in this condition monitoring of a gearbox (in IC Engine) is based on Machine Learning approach. Discrete frequencies were extracted from the sensor acquired signals, those signals were stored in as data at a sampling rate of 25.6kHz for all the load conditions with respective defects. (Refer Flowchart 1)

3.1 Data acquisition

The vibrational signals from the IC engine gearbox were measure using a tri-axial accelerometer. Different load conditions were manipulated using an Eddy current dynamometer. Totally there were 3 Load conditions (Load 1 – 9.6 Nm torque, Load 2 – 13.3 Nm torque and No Load) and 5 defect conditions (25% defect, 50% defect, 75% defect, 100% defect, No defect - refer Figure 2). The data acquisition process is explained in detail in section 3. The acquired signals need to be processed before the feature extraction process. The vibrational signals are then converted from analog to digital signals.

3.2 Feature extraction

The feature extraction was carried out in 3 ways

- Statistical features
- Histogram features
- Auto Regressive Moving Average (ARMA) features

Feature extraction is the process of converting the raw data obtained into numerical feature sets that can be computed easily and also by preserving the original information in the signal. The signal provided 25600 datapoints for each load and defect condition, it is then shortened into 14 statistical features, 100 histogram bins, 50 ARMA orders with 6-7 features to ease the computation process.

3.2.1 Statistical feature

The statistical feature extraction involves calculating various available statistical measures from the presented raw data to derive the set of features for further computation. The feature captures certain characteristics of the acquired data and make them suitable for machine learning.

The features include **Mean, Standard Error, Median, Mode, Standard Deviation, Sample Variance, Kurtosis, Skewness, Range, Minimum, Maximum, Sum**. These features provide insights into the underlying patterns and their relation with the acquired data.

3.2.2 Histogram feature

The histogram feature extraction involves the process of dividing a continuous variable data into intervals (also called as bins) and calculating the frequency of data points that is categorized in each interval. The maximum value is taken from the Statistical "maximum" feature and the minimum value is also taken from the Statistical "minimum" feature. The range is formed from the minimum and maximum, then the data points are separated in respective bins between the intervals 2-100. The best bins are chosen after which feature selection process happens through the J48 classifier and a lot of features are manually removed.

3.2.3 ARMA feature

Auto Regressive Moving Average (ARMA) feature extraction is a technique commonly used on quantitative data. ARMA models generally capture the underlying patterns and the relations in data by autoregressive (AR) and moving average (MA) components. These components model the data's dependency on its past errors and past values, respectively[22].

3.3 Feature selection

This study utilizes the power the decision tree classifier, namely J48. This classifier is adept and accurate in handling complex data points. The high accuracy features for each type are identified and these features undergo the data mining process to uncover the redundancy using the J48 algorithm. Manual validation and checking of the features help refine the Decision tree and drop the irrelevant features and keeping the relevant features without compromising with the accuracies and countering the potential over-fitting.

3.3.1 Statistical feature selection

For the Statistical features of each load conditions, the selection process had to be done only once. The data files were saved as CSV (Comma Separated Values) files for being recognized in WEKA software. The J48 tree classifier was used on the dataset and the accuracy was noted down.

The J48 decision tree was analyzed and a few features were selected manually using the decision tree. The features that were not selected in this process were removed from the dataset. The dataset now contains only the selected features, which were used for the feature classification. For each load condition, different statistical features were chosen. Pruning was then done to cut off more unimportant features in order to increase the accuracy (Refer Table 2). The selected statistical features and their respective load conditions are mentioned in the Table 3.

Table 2 - Statistical feature pruning

No load		Load 1		Load 2	
No of features	**Accuracy (%)**	**No of features**	**Accuracy (%)**	**No of features**	**Accuracy (%)**
6	79.33	**4**	**90.67**	5	90.00
5	79.33	3	86.00	4	90.67
4	78.67	2	83.00	**3**	**91.33**
3	**82.67**	1	80.00	2	88.67
2	72.67	0	20.00	1	77.33
1	53.33			0	20.00
0	20.00				

Table 3 - Statistical Selected features

LOAD CONDITIONS	SELECTED FEATURES
No Load	Kurtosis, Skewness, Range
Load 1 (Torque = 9.6 Nm)	Standard Error, Median, Kurtosis, Range
Load 2 (Torque = 13.3 Nm)	Median, Kurtosis, Range

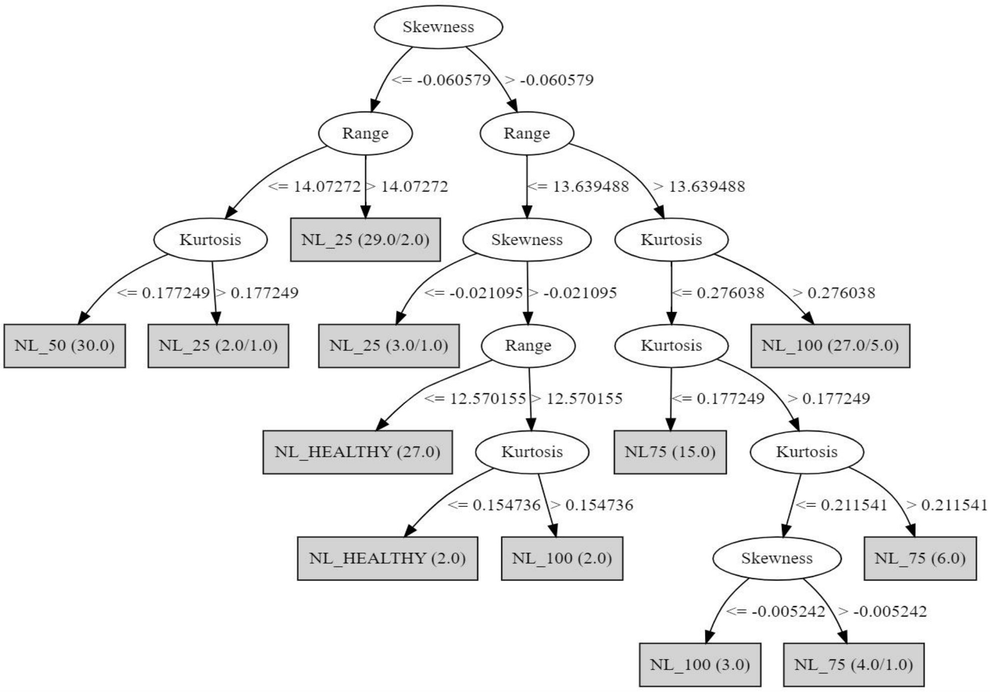


Figure 4 - Statistical – No Load – J48 Decision Tree

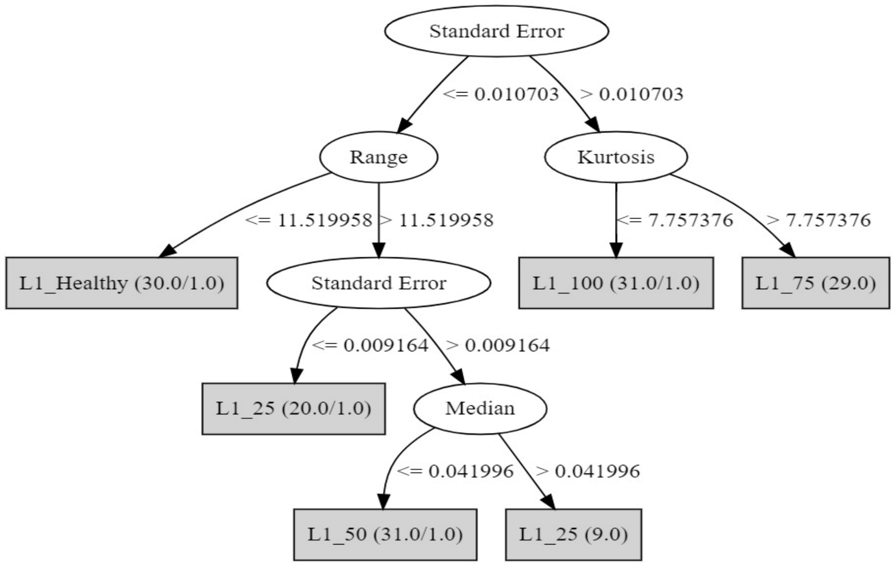


Figure 5 - Statistical – Load 1 – J48 Decision Tree

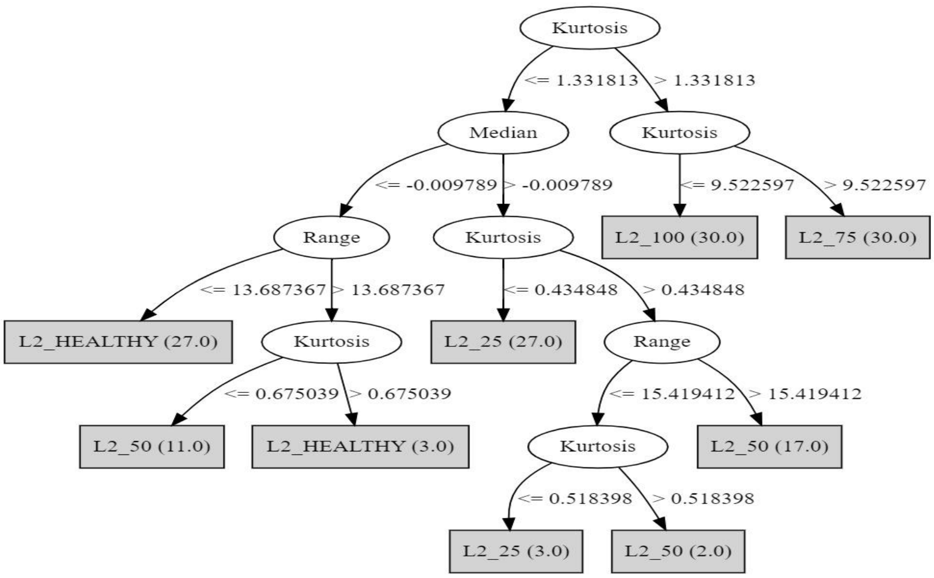


Figure 6 - Statistical – Load 2 – J48 Decision Tree

3.3.2 Histogram feature selection

The feature selection for the Histogram features are similar to that of the statistical. The data points are organized into bins ranging from 2 to 100 bins, the same process has to be followed as done as in statistical feature selection. Each bin needs to be divided into the 3 load conditions with the 5 fault conditions of the gearbox. Once this process has been completed, J48 decision tree algorithm is used. For each bin of each load conditions, the accuracy level is noted down. The best bins with the best accuracies for each load conditions are chosen, the J48 decision tree of those bins were analysed and the relevant (important) features were selected. The pruning process for the histogram features are mentioned in Table 4. Only the selected histogram features in each bin was considered for the feature classification step. The best bins are mentioned in Table 5 with their accuracy and the selected histogram bin features are listed for each load in Table 6.

Table 4 - Histogram feature pruning

No Load Bin67 (36,30,29,23,32,26)		**Load 1 Bin50 (14,15,21,23, 24)**		**Load 2 Bin 57 (37,27,24,20,12)**	
Overall	90.6667	**Overall**	97.3333	**Overall**	98
6 features	**87.3333**	**5 features**	**98**	5 features	96.6667
5 features	76	4 features	88	**4 features**	**96.6667**
4 features	68.6667	3 features	82.6667	3 features	82
3 features	63.3333	2 features	55.3333	2 features	76
2 features	45.3333	1 feature	51.3333	1 feature	44
1 feature	36				

From Figure 7, it can be observed that the 67th bin has an accuracy of 90.67 % which is the highest among all other bins for the No load condition. The Decision tree for the feature selection of the 67th bin is shown in Figure 8

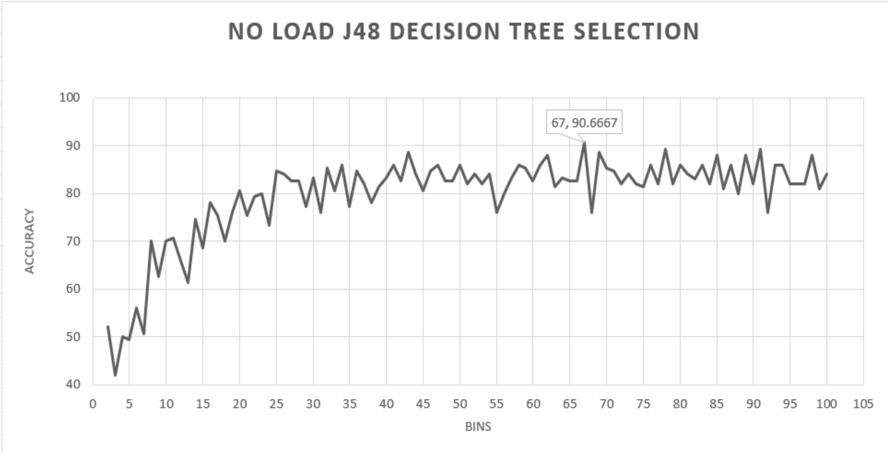


Figure 7 - Best Bin selection of Histogram features (No load)

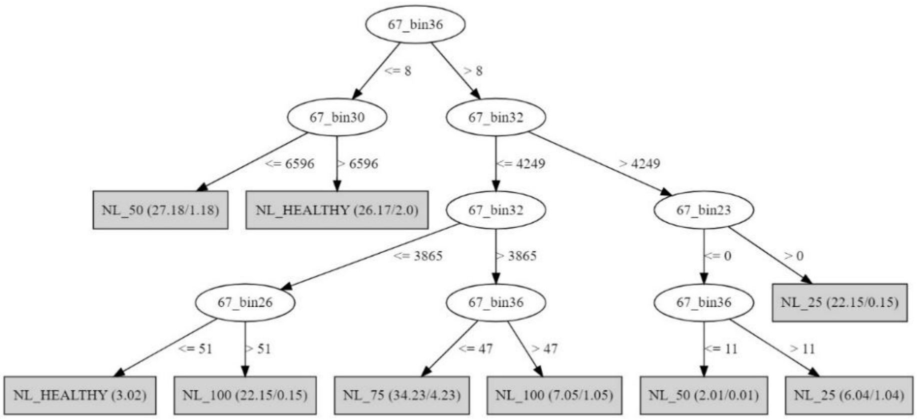


Figure 8 - Histogram Decision tree of 67th bin (No Load)

From Figure 9 it can be observed that the 50th bin has an accuracy of 97.33 % which is the highest among all other bins for the Load 1 condition. The Decision tree for the feature selection of the 50th bin is shown in Figure 10

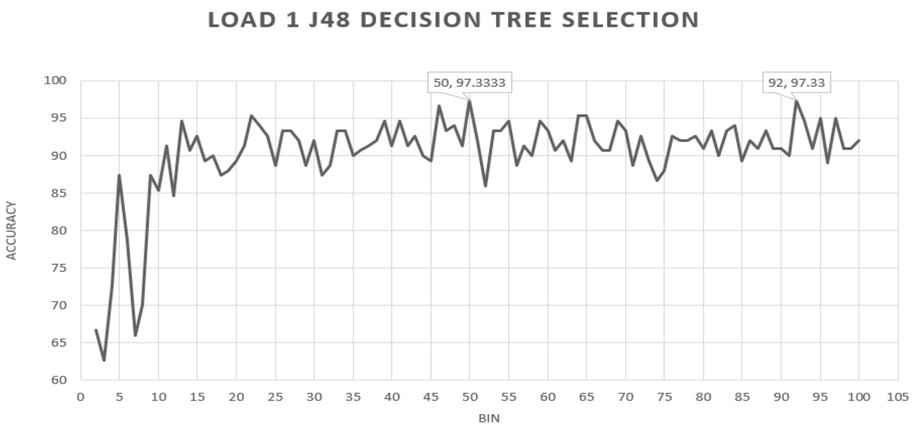


Figure 9 - Best Bin selection of Histogram features (Load 1)

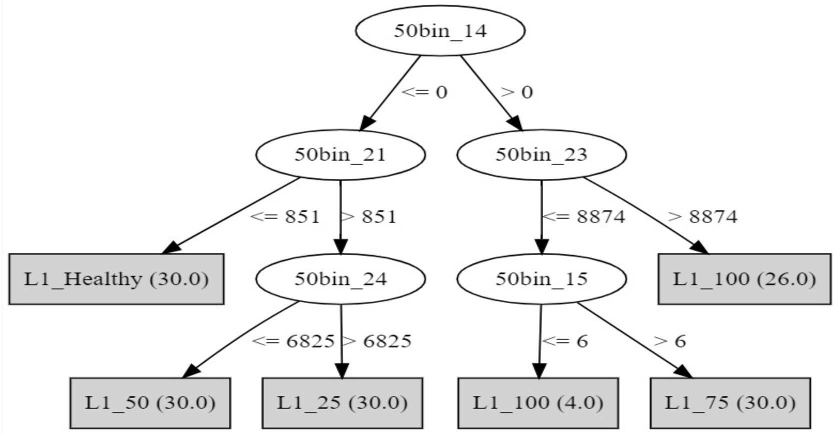


Figure 10 - Histogram Decision tree of 50th bin (Load 1)

From Figure 11 it can be observed that the 57th bin has an accuracy of 98 % which is the highest among all other bins for the Load 2 condition. The Decision tree for the feature selection of the 57th bin is shown in Figure 12

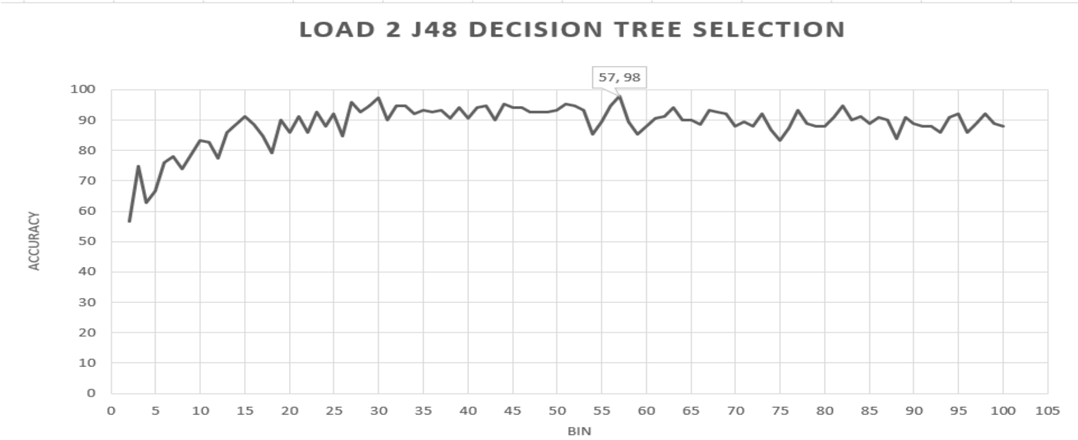


Figure 11 - Best Bin selection for Histogram features (Load 2)

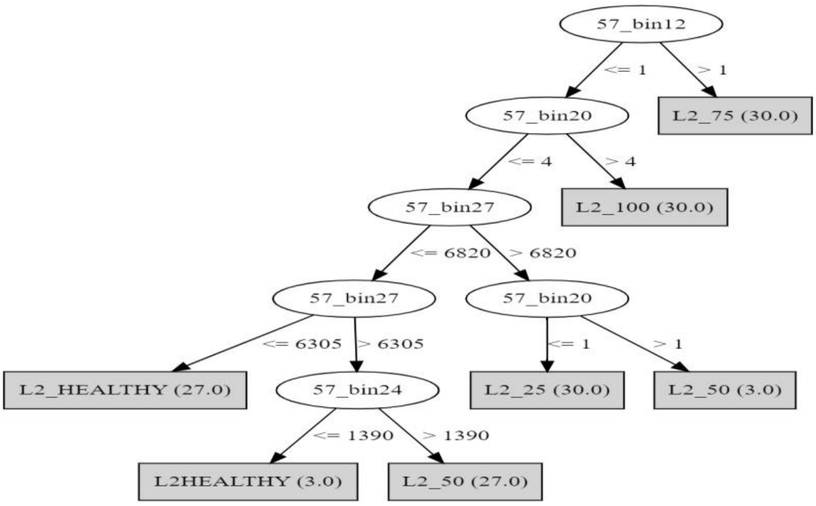


Figure 12 - Histogram Decision tree of 57th bin (Load 2)

Table 5 - Histogram Best Bins with their accuracy

LOAD CONDITIONS	BEST BINS	ACCURACY (%)
No Load	67	90.67
Load 1 (Torque = 9.6 Nm)	50	97.33
Load 2 (Torque = 13.3 Nm)	57	98

Table 6 - Selected Bin features

LOAD CONDITION	SELECTED FEATURES
NO LOAD	67_bin23, 67_bin26, 67_bin29, 67_bin30, 67_bin32, 67_bin36
LOAD 1	50bin_14, 50bin_15, 50bin_21, 50bin_23, 50bin_24
LOAD 2	57bin_12, 57bin_20, 57bin_24, 57bin_27

3.3.3 ARMA feature selection

Similar process was carried out for ARMA features as well, the accuracy for each order from the J48 decision tree was noted down and the best order's features were selected and then pruned. The selected feature are illustrated in Table 7 'a' refers to auto regression, 'k' refers to moving average and 'e' refers to error.

From Figure 13 it can be observed that the 50th order has an accuracy of 98 % which is the highest among all other orders for the No load condition. The Decision tree for the feature selection of the 50th order is shown in Figure 14

Figure 13 - ARMA feature selection graph (No load)

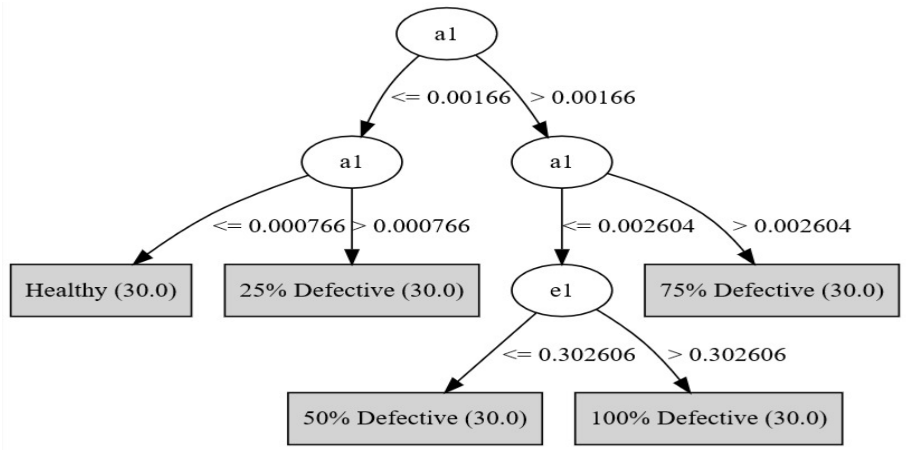


Figure 14 - ARMA Decision tree of 50th order (No Load)

From Figure 15 it can be observed that the 6th order has an accuracy of 98.67 % which is the highest among all other orders for the Load 1 condition. The Decision tree for the feature selection of the 6th order is shown in Figure 16

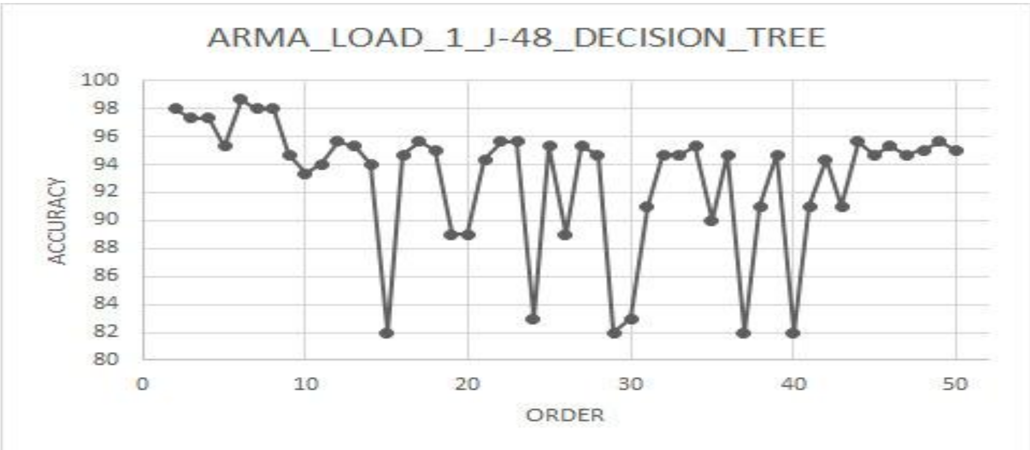


Figure 15 - ARMA feature selection graph (Load 1)

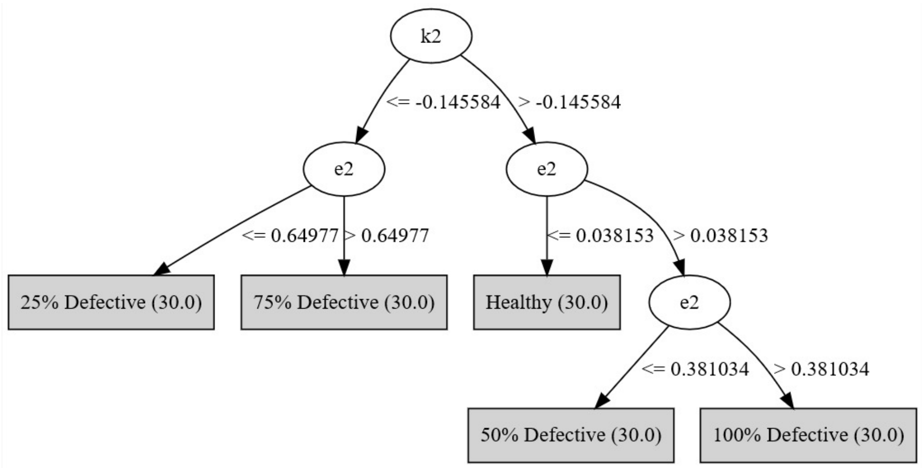


Figure 16 - ARMA Decision tree of 6th order (Load 1)

From Figure 17 it can be observed that the 2^{nd} order has an accuracy of 98.67 % which is the highest among all other orders for the Load 2 condition. The Decision tree for the feature selection of the 2^{nd} order is shown in Figure 18

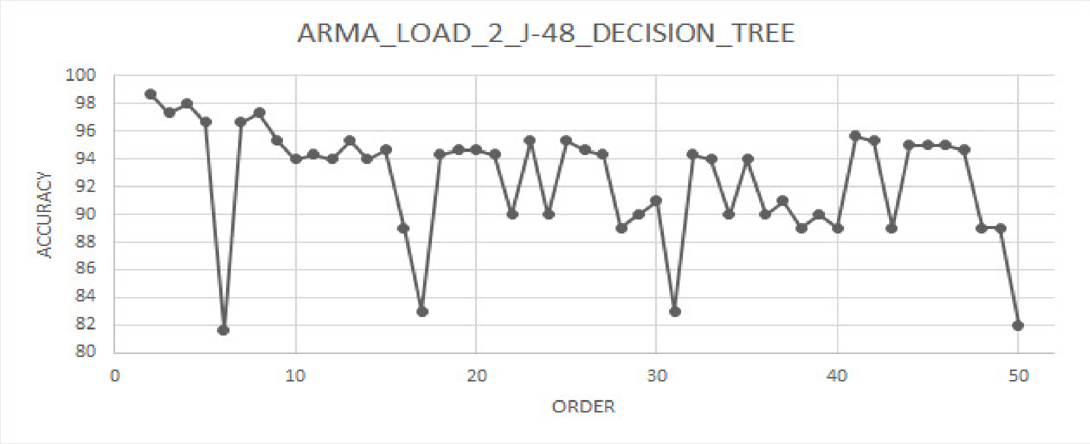


Figure 17 - ARMA Decision tree of 6^{th} order (Load 1)

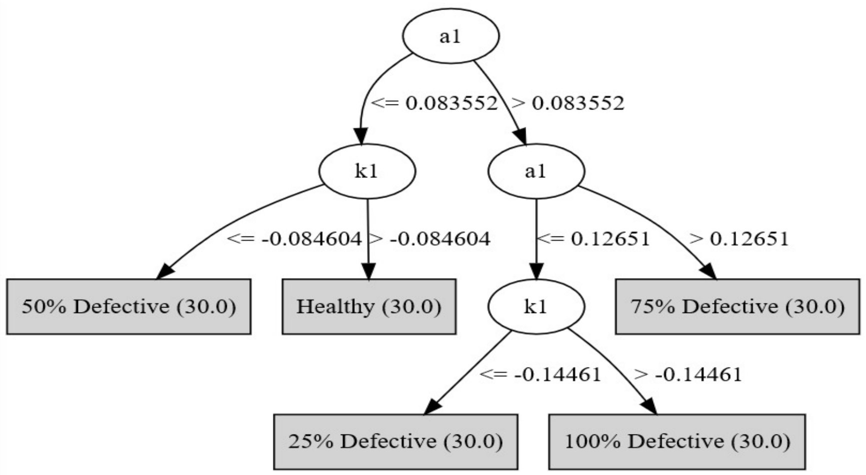


Figure 18 - ARMA Decision tree of 2^{nd} order (Load 2)

Table 7 - Selected ARMA features for each load condition

LOAD CONDITIONS	SELECTED FEATURES	ORDER	ACCURACY (%)
No Load	a1, e1	50	98
Load 1 (Torque = 9.6 Nm)	e2, k2	6	98.67
Load 2 (Torque = 13.3 Nm)	a1, k1	2	98.67

3.4 Feature Classification

8 Machine learning classifiers are chosen for the feature classification process. The selected data were first separated into Training dataset (80%) and Testing dataset (20%) and the classifier models were applied to find the Training accuracy, validation accuracy and Testing accuracy. All the values were noted down for selecting the classifier with best accuracy. Voting classifier was used and 2,3,4,5 ensemble models were formed with the classifiers in order to improve the accuracy wherever possible [13]. (The results obtained after using these 8 classifiers are shown in section 4)

3.4.1 Classifiers: Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised classification algorithm that finds a hyperplane to separate different classes. It aims to extend the margin between those classes. The SVM classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.2 Classifiers: Multi-layer perceptron (MLP)

Multilayer perceptron is a type of artificial neural network with multiple layers of interconnected nodes. It is capable of learning complex relations in datasets. The MLP classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down

3.4.3 Classifiers: Logistics (LO)

Logistics Regression classifier is a linear classification algorithm that models the probability of binary outcomes. The major goal is to predict one of two possible classes. It can also be used for certain multiclass classification using certain techniques. The LO classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.4 Classifiers: Random Forest (RF)

Random Forest is a powerful machine learning algorithm used in classification (and regression as well). Random Forest is known for its versatility, robustness and the ability to handle complex data. Random Forest is a classifier that contains a number of decision trees on various other subsets of the provided dataset and takes the average to improve the model's predictive accuracy[23]. Greater the number of trees in the forest leads to higher accuracy and helps in preventing the problem of overfitting.

The RF classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.5 Classifiers: J48

J48 is a Decision tree classification algorithm, used for classification and regression. It constructs decision trees from training data for the classification task by recursively splitting the data based on the selected features and then create branches and combine them to form a classification outcome. The J48 classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.6 Classifiers: Logistic Model Tree (LMT)

The Logistic Model Tree (LMT) classifier is a machine learning algorithm that combines the decision tree structures with logistic regression models. It aims to capture both non-linear and linear relationships in the data points by leveraging the strength of decision tree and logistic regression. It is useful when the relationship between the class labels and the features are complex. The LMT classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.7 Classifiers: Naïve Bayes (NB)

Naïve Bayes is a supervised machine learning algorithm which is based on the Bayes theorem and it helps in classification problems. It is one of the simple and most effective classifier algorithms which helps us in building learning models that can make quick predictions [24]. It is a probabilistic classifier, meaning it predicts on the basis of the probability of the outcome. It is called Naïve because it assumes that the occurrence of certain features is independent of the occurrence of other features. The NB classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down.

3.4.8 Classifiers: k – Nearest Neighbor (k-NN)

The k – Nearest Neighbor is a machine learning algorithm that is based on supervised learning. The k-NN algorithm assumes the similarity between the new available data and put up the new data in the category that is more similar to the available dataset categories. It classifies the new datapoints (Test) based on similarity [26]. The k-NN classifier was applied for the selected features of each load conditions and the Training accuracy, Validation accuracy and the Test accuracy was noted down. The table below shows the accuracies of the k-NN classifier for each feature and their respective loads.

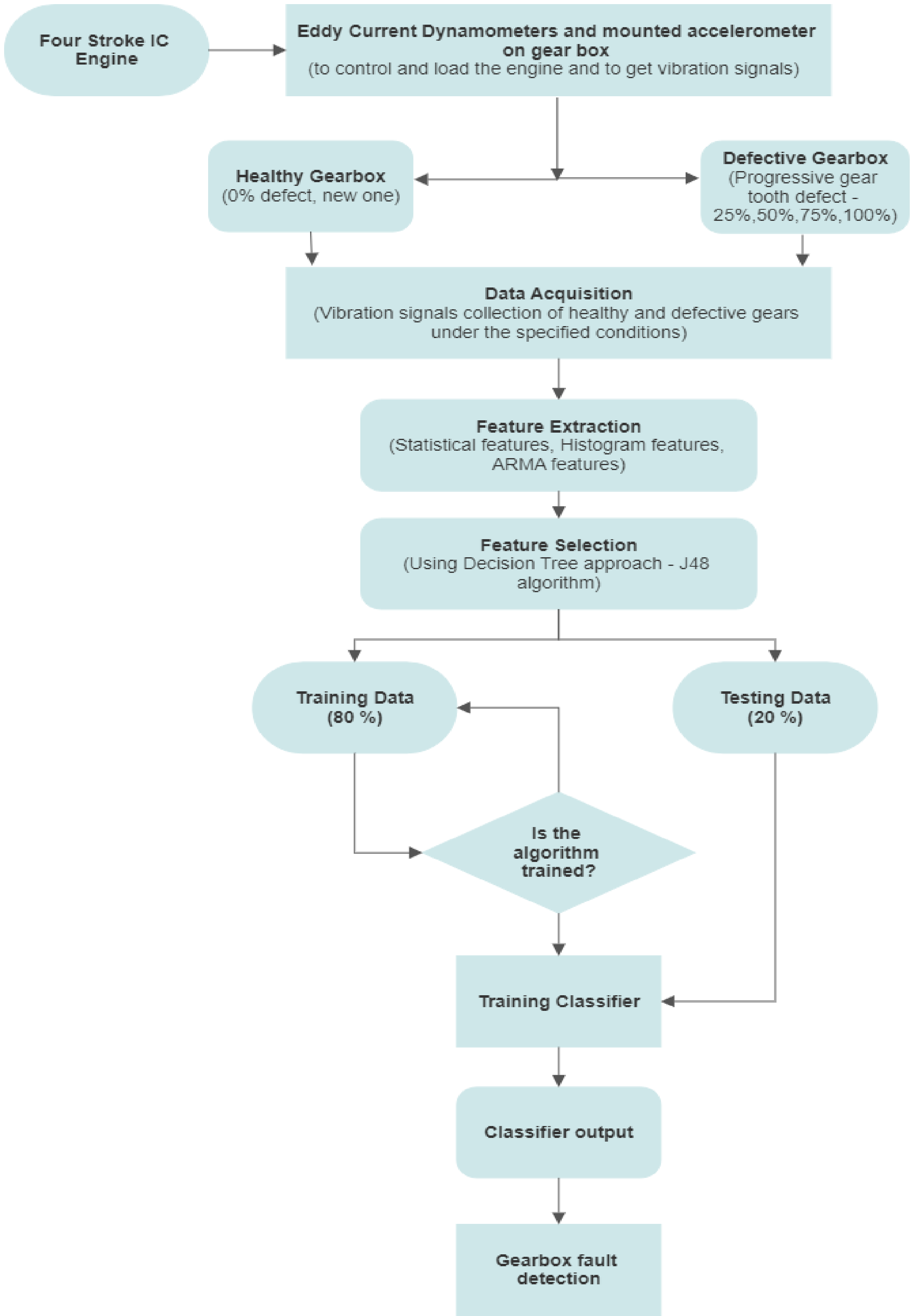
3.5 Evaluation metrics

The evaluation of the classifier performance is done based on 3 metrics: Training accuracy, Validation accuracy and Testing accuracy.

Training accuracy measures how well the model performs on the train data, a high training accuracy indicates that the classifier model has learned the patterns in the training set very well. However, a very high training accuracy may also suggest overfitting of the data, where the model has memorized the training data points and it may not generalize well to the new data.

Validation accuracy assesses the classifier model's performance on a separate validation set, but in our case, it was cross validation where, the training data was divided into 10 folds and then validated separately. The validation set provides how well the model is expected to perform on new, unseen data.

Testing accuracy measures the performance of the classifier model on an entirely new and unseen test dataset. Testing accuracy is the final evaluation of the model's generalization performance and also provides an estimate of how the model is likely to perform in real world scenarios when faced with real problems. Better the testing accuracy, better is the fault detection and diagnosis of the IC Engine Gearbox



Flowchart – 1: Methodology

4. Results and Discussion

From the experimental testing done, the various fault conditions were induced and the vibrational readings were taken from each of the Load conditions. A total of 5 faults in each load condition. Feature extraction was done and the datasets were separated. After they had been separated into different datasets, feature selection process was carried out using the Decision Tree classifier – J48 algorithm. And the results for each were noted down

4.1 Feature classification

The 8 classifiers were run on each of the selected features for the different load conditions. The accuracies were noted down. Some of the Histogram and ARMA features resulted in 100% accuracies after the feature classification. For statistical features, there was no 100% accuracy in each of the three load conditions. The time taken to create a model for testing, training and validation is very important, since most of these classifiers are powerful in computation. To choose the best classifier out of the ones that produce 100% accuracy, we would need the time consumed in building those classifier models. The time was noted only for those features where the accuracy was deemed to be 100% The wear and tear of the Gearbox can be identified from the results of the best models. Voting classifier was used to try and improve the accuracies of the Statistical feature classification. Load 1 (9.6 Nm Torque) and Load 2 (13.3 Nm Torque) did not show any improvement in the Test accuracy level, but the Load 2 showed improvements in Validation accuracy. Whereas, No load condition showed approximately 4% improvement in the Test accuracy for a certain ensemble voting model.

A voting classifier is an ensemble machine learning model used to combine the results of multiple individual classifiers to make a final accuracy prediction that generally gives a better value than the individual classifiers. It has different combination rules as listed below:

- Average of Probabilities
- Product of Probabilities
- Majority Voting
- Minimum Probability
- Maximum Probability

The classifier combinations were made on the top 5 classifiers out of the 8. 2,3,4,5 ensemble models are separated and classified. The Performance of the two class ensemble model, three class ensemble model, four class ensemble model, five class ensemble models are noted and the best Test accuracy resulted from the voting classification is noted separately for the proper fault detection of the gearbox.

4.1.1 Classifiers: Support Vector Machine (SVM)

Table 8 - SVM

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	97.50	90.00	93.33
ARMA_L2	98.33	97.50	96.67
ARMA_NL	96.67	67.50	93.33
HISTO_L1	98.33	99.17	96.67
HISTO_L2	86.67	86.67	93.33
HISTO_NL	75.83	70.00	73.33
STAT_L1	90.83	78.33	86.67
STAT_L2	83.33	82.50	83.33
STAT_NL	86.67	82.50	76.67

4.1.2 Classifiers: Multi-layer perceptron (MLP)

Table 9 - MLP

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	100.00
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	100.00	100.00
HISTO_L1	100.00	100.00	100.00
HISTO_L2	96.67	95.00	100.00
HISTO_NL	97.50	92.50	100.00
STAT_L1	100.00	95.00	96.67
STAT_L2	98.33	95.00	93.33
STAT_NL	96.67	93.33	73.33

4.1.3 Classifiers: Logistics (LO)

Table 10 - LO

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	96.67
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	99.00	100.00
HISTO_L1	100.00	98.33	100.00
HISTO_L2	100.00	92.50	96.67
HISTO_NL	97.50	90.00	86.67
STAT_L1	100.00	95.00	90.00
STAT_L2	100.00	95.00	93.33
STAT_NL	98.33	88.33	80.00

4.1.4 Classifiers: Random Forest (RF)

Table 11 - RF

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	93.33
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	100.00	100.00
HISTO_L1	100.00	96.67	93.33
HISTO_L2	100.00	97.50	100.00
HISTO_NL	100.00	92.50	100.00
STAT_L1	100.00	95.83	86.67
STAT_L2	100.00	92.50	93.33
STAT_NL	100.00	83.33	86.67

4.1.5 Classifiers: J48

Table 12 - J48

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	98.33	93.33
ARMA_L2	100.00	96.67	96.67
ARMA_NL	100.00	97.50	93.33
HISTO_L1	100.00	96.67	86.67
HISTO_L2	100.00	96.67	96.67
HISTO_NL	98.33	85.83	90.00
STAT_L1	98.33	94.17	86.67
STAT_L2	100.00	87.50	96.67
STAT_NL	95.00	85.33	76.67

4.1.6 Classifiers: Logistic Model Tree (LMT)

Table 13 - LMT

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	93.33
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	99.17	100.00
HISTO_L1	100.00	99.17	96.67
HISTO_L2	99.17	97.50	100.00
HISTO_NL	95.00	90.83	93.33
STAT_L1	98.33	96.67	93.33
STAT_L2	95.00	90.00	80.00
STAT_NL	95.00	90.00	80.00

4.1.7 Classifiers: Naïve Bayes (NB)

Table 14 - NB

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	100.00
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	100.00	100.00
HISTO_L1	97.50	96.67	93.33
HISTO_L2	98.33	98.33	100.00
HISTO_NL	91.67	89.17	93.33
STAT_L1	96.67	95.83	83.33
STAT_L2	95.83	94.17	93.33
STAT_NL	91.67	87.50	83.33

4.1.8 Classifiers: k – Nearest Neighbor (k-NN)

Table 15 - k-NN

FEATURES	TRAINING ACCURACY (%)	CROSS VALIDATION (%)	TEST ACCURACY (%)
ARMA_L1	100.00	100.00	96.67
ARMA_L2	100.00	100.00	100.00
ARMA_NL	100.00	100.00	100.00
HISTO_L1	100.00	100.00	100.00
HISTO_L2	100.00	96.67	100.00
HISTO_NL	100.00	90.83	100.00
STAT_L1	100.00	99.17	93.33
STAT_L2	100.00	94.17	96.67
STAT_NL	100.00	89.17	76.67

Histogram and ARMA feature classification resulted in 100% Test accuracy in most of the classifiers for each load condition and as mentioned above, any of the classifiers with the 100% accuracy can be chosen for this process to choose the absolute best, we have to look at the time taken for the Training, Validation and Testing model to be formed for classification. The best classifier would result in faster computation of the fault detection and diagnosis of the IC engine gearbox with a greater validation. The statistical feature classification results for each load has been discussed in the below section.

4.2 Statistical Feature Classification of each load

Table 16 – Load 1

S. NO	CLASSIFIER	TRAINING ACCURACY (%)	VALIDATION ACCURACY (%)	TEST ACCURACY (%)
1	MLP	100.00	95.00	96.67
2	RF	100.00	95.83	86.67
3	NB	96.67	95.83	83.33
4	LMT	98.33	96.67	93.33
5	J48	98.33	94.17	86.67
6	LO	100.00	95.00	90.00
7	SVM	90.83	78.33	86.67
8	k-NN	100.00	99.17	93.33

Table 17 - Load 2

S. NO	CLASSIFIER	TRAINING ACCURACY (%)	VALIDATION ACCURACY (%)	TEST ACCURACY (%)
1	MLP	98.33	95.00	93.33
2	RF	100.00	92.50	93.33
3	NB	95.83	94.17	93.33
4	LMT	99.17	96.67	93.33
5	J48	100.00	87.50	96.67
6	LO	100.00	95.00	93.33
7	SVM	83.33	82.50	83.33
8	k-NN	100.00	94.17	96.67

Table 18 – No Load

S. NO	CLASSIFIER	TRAINING ACCURACY (%)	VALIDATION ACCURACY (%)	TEST ACCURACY (%)
1	MLP	96.67	93.33	73.33
2	RF	100.00	83.33	86.67
3	NB	91.67	87.50	83.33
4	LMT	95.00	90.00	80.00
5	J48	95.00	85.33	76.67
6	LO	98.33	88.33	80.00
7	SVM	86.67	82.50	76.67
8	k-NN	100.00	89.17	76.67

4.3 Voting classifier

The statistical feature selection process did not result in 100% accuracy from any of the classifiers. So, voting was performed on the Statistical features (Load 1, Load 2, No Load). Load 1 and No Load did not show any improvement in the Test accuracy after Voting. The Load 2 statistical features upon performing voting classifiers, showed an improvement in the Testing accuracy for certain combination. The best results are shown in Table 19

The top 5 classifiers for the Load 2 Statistical features were (Refer Table 17):

- K-Nearest Neighbor (KNN)
- J48
- Logistic Model Tree (LMT)
- Logistics (LO)
- Multi-layer Perceptron (MLP)

When used separately, the classifiers provide a maximum accuracy of 96.67 % for the Load statistical features. When voting was applied, some combinations of classifiers produced a 100 % test accuracy, which is approximately 3.33 % greater than the test result from individual classifiers. The Best Voting combinations that resulted in better obtained are mentioned below:

- 2 class ensemble model:
  - (J48 – LO)
- 3 class ensemble model:
  - (KNN – J48 - LMT)
  - (KNN – J48 - LO)
  - (KNN – J48 – MLP)
  - (J48 – LMT – LO)
  - (J48 – LMT – MLP)
  - (J48 – LO – MLP)
- 4 class ensemble model:
  - (KNN – J48 – LMT – LO)
  - (KNN – J48 – LMT – MLP)
  - (KNN – J48 – LO – MLP)
  - (J48 – LMT – LO – MLP)
- 4 class ensemble model:
  - (KNN – J48 – LMT – LO - MLP)

Table 19 - Performance of two class ensemble model (J48 – LO)

S.No	Voting Strategy	Training Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)
1	Average of probabilities (AOP)	100	89.17	**100**
2	Product of probabilities (POP)	100	86.67	96.67
3	Majority Voting (MV)	100	89.17	**100**
4	Minimum probability (MINP)	100	86.67	96.67
5	Maximum probability (MAXP)	100	89.17	**100**

2 class ensemble models are preferred over the rest because, their computation is simpler and more efficient than other multi-class models.

TARGET / OUTPUT	L225	L250	L275	L2100	L2HEALTHY	SUM
L225	6 20.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	6 100.00% 0.00%
L250	0 0.00%	6 20.00%	0 0.00%	0 0.00%	0 0.00%	6 100.00% 0.00%
L275	0 0.00%	0 0.00%	6 20.00%	0 0.00%	0 0.00%	6 100.00% 0.00%
L2100	0 0.00%	0 0.00%	0 0.00%	6 20.00%	0 0.00%	6 100.00% 0.00%
L2HEALTHY	0 0.00%	0 0.00%	0 0.00%	0 0.00%	6 20.00%	6 100.00% 0.00%
SUM	6 100.00% 0.00%	6 100.00% 0.00%	6 100.00% 0.00%	6 100.00% 0.00%	6 100.00% 0.00%	30 / 30 100.00% 0.00%

Figure 19 – Confusion matrix for AOP, MV, MAXP (J48-LO)

5 Conclusion

The improvement in accuracy realized by combining the classifier models under the majority of operational settings and opens up intriguing direction for further study and development. Our study has shown the power of feature engineering and feature classification in the fault detection and diagnosis. There are several areas where these methods can be beneficial.

There is tremendous potential in developing real-time adaptation mechanisms, where the system dynamically modifies its feature engineering, classification and problem-solving techniques in response to shifting operational conditions. As a result, an autonomous and adaptable diagnostic system would be created that could continuously deliver accurate results regardless of changes in load or the operational environment or even any defects.

In our work, we had set out on a thorough analysis of how feature engineering, and feature classification using the individual machine learning classifiers and a combination of those classifiers for some cases to improve the Testing accuracy of the model. The machine learning models for the gearbox fault diagnosis represents a significant advancement in the field of predictive maintenance of the IC engine components.

The incorporation of voting classifiers added an additional layer of sturdiness and the accuracy of the diagnosis process. An accuracy of **100%** was reached for Load 2 condition for 2,3,4,5 class ensemble models and the 2-class model were chosen to the best for easier computation. In essence, this study underlines the importance of a comprehensive and structured approach to gearbox fault diagnosis. As industries continue to prioritize the operational efficiency and downtime reduction, the insights and machine learning methodologies presented in this paper hold he potential to significantly contribute to optimization of predictive maintenance strategies and the overall reliability of the machinery.

REFERENCES:

[1] R. KN, H. Kumar, K. GN, and G. KV, “Fault diagnosis of internal combustion engine gearbox using vibration signals based on signal processing techniques,” *J Qual Maint Eng*, vol. 27, no. 2, pp. 385–412, 2020, doi: 10.1108/JQME-11-2019-0109.

[2] K. N. Ravikumar, C. K. Madhusudana, H. Kumar, and K. V. Gangadharan, “Classification of gear faults in internal combustion (IC) engine gearbox using discrete wavelet transform features and K star algorithm,” *Engineering Science and Technology, an International Journal*, vol. 30, Jun. 2022, doi: 10.1016/j.jestch.2021.08.005.

[3] X. Liang, M. J. Zuo, and Z. Feng, “Dynamic modeling of gearbox faults: A review,” *Mechanical Systems and Signal Processing*, vol. 98. Academic Press, pp. 852–876, Jan. 01, 2018. doi: 10.1016/j.ymssp.2017.05.024.

[4] C. Li and M. Liang, “Extraction of oil debris signature using integral enhanced empirical mode decomposition and correlated reconstruction,” *Meas Sci Technol*, vol. 22, no. 8, 2011, doi: 10.1088/0957-0233/22/8/085701.

[5] A. M. D. Younus and B. S. Yang, “Intelligent fault diagnosis of rotating machinery using infrared thermal image,” in *Expert Systems with Applications*, Feb. 2012, pp. 2082–2091. doi: 10.1016/j.eswa.2011.08.004.

[6] C. K. Madhusudana, H. Kumar, and S. Narendranath, “Face milling tool condition monitoring using sound signal,” *International Journal of System Assurance Engineering and Management*, vol. 8, pp. 1643–1653, Nov. 2017, doi: 10.1007/s13198-017-0637-1.

[7] Z. Liu, X. Wang, and L. Zhang, “Fault Diagnosis of Industrial Wind Turbine Blade Bearing Using Acoustic Emission Analysis,” *IEEE Trans Instrum Meas*, vol. 69, no. 9, pp. 6630–6639, Sep. 2020, doi: 10.1109/TIM.2020.2969062.

[8] V. Indira, R. Vasanthakumari, R. Jegadeeshwaran, and V. Sugumaran, “Determination of minimum sample size for fault diagnosis of automobile hydraulic brake system using power analysis,” *Engineering Science and Technology, an International Journal*, vol. 18, no. 1, pp. 59–69, Mar. 2015, doi: 10.1016/j.jestch.2014.09.007.

[9] W. Teng, X. Ding, S. Tang, J. Xu, B. Shi, and Y. Liu, “Vibration analysis for fault detection of wind turbine drivetrains—a comprehensive investigation,” *Sensors*, vol. 21, no. 5. MDPI AG, pp. 1–33, Mar. 02, 2021. doi: 10.3390/s21051686.

[10] M. Vishwakarma, R. Purohit, V. Harshlata, and P. Rajput, “Vibration Analysis & Condition Monitoring for Rotating Machines: A Review,” in *Materials Today: Proceedings*, Elsevier Ltd, 2017, pp. 2659–2664. doi: 10.1016/j.matpr.2017.02.140.

[11] K. Vernekar, H. Kumar, and K. V. Gangadharan, “Engine gearbox fault diagnosis using machine learning approach,” *J Qual Maint Eng*, vol. 24, no. 3, pp. 345–357, Aug. 2018, doi: 10.1108/JQME-11-2015-0058.

[12] F. Al-Badour, M. Sunar, and L. Cheded, “Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques,” *Mech Syst Signal Process*, vol. 25, no. 6, pp. 2083–2101, Aug. 2011, doi: 10.1016/j.ymssp.2011.01.017.

[13] A. Vinayagam, V. Veerasamy, P. Radhakrishnan, M. Sepperumal, and K. Ramaiyan, “An ensemble approach of classification model for detection and classification of power quality

disturbances in PV integrated microgrid network," *Appl Soft Comput*, vol. 106, Jul. 2021, doi: 10.1016/j.asoc.2021.107294.

[14] J. Shi, D. Peng, Z. Peng, Z. Zhang, K. Goebel, and D. Wu, "Planetary gearbox fault diagnosis using bidirectional-convolutional LSTM networks," *Mech Syst Signal Process*, vol. 162, Jan. 2022, doi: 10.1016/j.ymssp.2021.107996.

[15] R. Ade, "Feature Selection based Classification using Naive Bayes, J48 and Support Vector Machine Dipali Bhosale," 2014.

[16] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, and Z. Zhang, "Compound Fault Diagnosis of Gearboxes via Multi-label Convolutional Neural Network and Wavelet Transform," *Comput Ind*, vol. 113, Dec. 2019, doi: 10.1016/j.compind.2019.103132.

[17] V. Sugumaran and K. I. Ramachandran, "Effect of number of features on classification of roller bearing faults using SVM and PSVM," *Expert Syst Appl*, vol. 38, no. 4, pp. 4088–4096, Apr. 2011, doi: 10.1016/j.eswa.2010.09.072.

[18] A. Moosavian, G. Najafi, B. Ghobadian, M. Mirsalim, S. M. Jafari, and P. Sharghi, "Piston scuffing fault and its identification in an IC engine by vibration analysis," *Applied Acoustics*, vol. 102, pp. 40–48, Jan. 2016, doi: 10.1016/j.apacoust.2015.09.002.

[19] J. Wang, S. Li, Y. Xin, and Z. An, "Gear Fault Intelligent Diagnosis Based on Frequency-Domain Feature Extraction," *Journal of Vibration Engineering and Technologies*, vol. 7, no. 2, pp. 159–166, Apr. 2019, doi: 10.1007/s42417-019-00089-1.

[20] F. Ye, Z. Zhang, K. Chakrabarty, and X. Gu, "Board-level functional fault diagnosis using artificial neural networks, support-vector machines, and weighted-majority voting," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 5, pp. 723–736, 2013, doi: 10.1109/TCAD.2012.2234827.

[21] S. Shaul Hameed, V. Muralidharan, and B. K. Ane, "Comparative analysis of fuzzy classifier and ANN with histogram features for defect detection and classification in planetary gearbox," *Appl Soft Comput*, vol. 106, Jul. 2021, doi: 10.1016/j.asoc.2021.107306.

[22] A. Joshuva and V. Sugumaran, "A machine learning approach for condition monitoring of wind turbine blade using autoregressive moving average (ARMA) features through vibration signals: a comparative study."

[23] S. Aravinth and V. Sugumaran, "Air compressor fault diagnosis through statistical feature extraction and random forest classifier," 2018.

[24] P. G. Sreenath, G. Praveen Kumare, S. Pravin, K. N. Vikram, and M. Saimurugan, "Automobile Gearbox Fault Diagnosis Using Naive Bayes and Decision Tree Algorithm," *Applied Mechanics and Materials*, vol. 813–814, pp. 943–948, Nov. 2015, doi: 10.4028/www.scientific.net/amm.813-814.943.

[25] A. Joshuva and V. Sugumaran, "A lazy learning approach for condition monitoring of wind turbine blade using vibration signals and histogram features," *Measurement (Lond)*, vol. 152, Feb. 2020, doi: 10.1016/j.measurement.2019.107295.

[26] L. Jiang, Z. Cai, D. Wang, and S. Jiang, "Survey of Improving K-Nearest-Neighbor for Classification."