ORIGINAL PAPER

Multiclass Fault Classification of an Induction Motor Bearing Vibration Data Using Wavelet Packet Transform Features and Artificial Intelligence

Shilpi Yadav[1] · Raj Kumar Patel[2] · Vijay Pratap Singh[2]

Abstract

Purpose The bearing that supports the rotor shaft is one of the essential aspects of any spinning machine, particularly in induction motors. Maintaining the bearing condition with some degree of assurance is essential.

Methods In the present research, a two-level Wavelet Packet transform (WPT) has been employed for the filtration to investigate the meaningful vibration signal. An ANOVA F test and Mutual information method have been used for feature selection. The Logistic Regression (LR) and Support Vector Classifier (SVC) has been considered to classify the fault.

Results Eleven statistical features from each of the original signals and wavelet decomposed signals were calculated. The present work investigates the existence of a fault, the type of fault, and its severity. The LR and SVC Model are used to evaluate the performance of the optimum feature set obtained from feature selection techniques.

Conclusion The sub-band signal DD2 with SVC gives the best results for all three cases from the results obtained with the full set feature as compared to the LR technique. The grid search method along with SVC produced the greatest results, with three features provided. Thus the classification accuracy of 100 percent was achieved with only three features in the case of two classes, a maximum accuracy of 96.3% was obtained from four classes with optimal feature 8, and an accuracy of 94.6% with optimal 8 features for 10 class problems. Thus the present technique for bearing fault diagnosis can be implemented for practical purposes.

Keywords ANOVA F test · Logistic regression · Mutual information · Wavelet packet transform · Support vector classifier

Introduction

An induction motor is one of the main components on which production depends [1]. Safety and proper maintenance are necessary to avoid sudden production stoppage and financial loss. Bearing defect identification and perception of seriousness are critical such that proactive measures can be taken considerably sooner, and severe bearing and system failure can be prevented. Typically, flaws are caused by changes in speed and load, which can lead to early bearing failure. It is also crucial to align the bearing with the rest of the system. Unwanted noise and vibrations are produced by an unbalanced system, leading to bearing damage. Rolling surface wear is initiated and spread in the dirt, dust, and other foreign particles. Deep scratches, dents, and other faults might occur during a bearing setup if it is mounted incorrectly. During operation, these faults worsen and impact the bearing's performance. Bearing faults is one of the most occurring faults in a rotating machine, which needs to be taken care of at regular intervals. By monitoring the machine body's vibration, the aberrant bearing behavior can be seen. Vibration signature consists of various frequencies due to damage in any bearing part. During operation, the vibration from the machine's body hides the same generated frequency due to surface damage. As a result, the presence of a defect in any section of the bearing component necessitates a thorough examination of the vibration signal to extract necessary information.

Numerous authors have investigated a range of defect diagnostics approaches for rolling element bearings. When faults of varying severity levels happen in a single section, they all occur with a similar characteristic frequency,

✉ Shilpi Yadav
  shilpiyadav8@gmail.com

1 Rajkiya Engineering College, Kannauj, India

2 Rajkiya Engineering College, Sonbhadra, India

making fault severity estimate more difficult. Researchers have looked into various defect diagnostics approaches for rolling element bearings [2]. The application of artificial intelligence methods like support vector machine (SVM), fuzzy logic, artificial neural network (ANN), and others in bearing defect diagnosis has also been documented in the literature. Features can be extracted from the obtained vibration signals to train a classifier. The most widely employed statistical parameters are the kurtosis, root mean square (RMS), the average magnitude of the faulty frequency, and crest factor. For fault diagnosis (FD), the researchers used a decision tree approach to identify the best features.

Samanta et al. looked into the behavior of SVM and ANN in detecting gear faults [3]. Kankar et al. Faults used features acquired from time-domain signals in bearing components that have been classified using artificial intelligence approaches such as SVM and ANN [4]. Kankar et al. also recommended response surface methodology (RSM) investigates the outcomes of faults in various bearing elements on the system's stability for rotor-bearing [5]. Predicting the degree of defects in bearings is still challenging to work. Jiang et al. carried out an observational study to determine the severity of rotating equipment defects. The vibration signals for multiple frequency band energies (MFBE) are extracted for feature selection and statistical and residue signals are employed to estimate fault severity. In addition to identifying the damaged bearing component, measuring the bearing diagnostic also entails estimating the fault's severity. The current study aims to classify a fault, type of fault, and fault severity levels in every induction motor bearing. Faults of varying intensity levels in the same element have the same frequency of occurrence. As a result, it is difficult to classify bearing problems of varying severity levels. For the various bearing situations, eleven features are estimated in this work. Furthermore, features are chosen based on how responsive they are to the defects. Machine learning algorithms such as LR and SVC use these features as input. Various attribute filters are implemented and compared to select suitable attributes. The classification effectiveness of SVC and LR is compared using distinct information filters.

Data Description and Feature Extraction

From the literature, it has been found that the most suited signal for the investigation of a mechanical fault in an induction motor is a vibration signal whose amplitude is the function of time. Vibration sensors are required to record the vibration signal of the machine. The vibration sensor's placement plays a vital role in recording the exact signal with preciseness in data acquisition. With the help of vibration sensors, the mechanical vibration from the structure is converted into an electrical signal named vibration signal,

which consists of information of vibration parameters. The data for the present work were retrieved from the data center "Case Western Reserve University (CWRU)," which is open source [6]. The data are recorded vibration signal at various speeds and at different fault size conditions from bearing installed in a three-phase induction motor. The speed varies from 1797 to 1720 RPM. As the load increases, the speed decreases. At three distinct defect sizes, i.e., 0.007 inches, 0.014 inches, and 0.021 inches each with the no-load, one HP, two HP, and three HP load, the signal recording performs. These faults are terms as bearing's BFs, ORFs, and IRFs. A fault case considers three possible types of fault, three fault sizes, and four different loading conditions. Thus a total of 36 different fault conditions and four healthy condition data have been considered to classify the healthy and faulty condition of the bearing reported in Table 1. As per Table 1, 680 signals in which intact signals consist of 320 segments and fault signals consist of 360 segments. In vibration signal to analyze slight variations, each database has 3000 samples with a sampling frequency of 12,000 samples per second.

For classification purposes, a feature extraction is an important act. Once the vibration signal is obtained, the first action is to calculate the statistical parameters to analyze the time-domain signal and compare it with the baseline signal. If the amplitude of the statistical parameter changes significantly, it must be immediately taken care of before the major destruction occurs in an induction motor. The statistical parameters give the prior mechanical status regarding intact and faulty conditions. The popular statistical parameters which are extensively used for analyze the bearing defects are max (F1), min (F2), RMS (F3), mean (F4), mode (F5), standard deviation (F6), median (F7), variance (F8), skewness (F9), kurtosis (10), and energy of the signal (F11) [7].

After feature calculation, selecting the relevant feature that contributes most to the prediction variable or desire output is more important. The goal of feature selection is to exclude non-informative or unnecessary variables from the model. Having too much irrelevant data leads to a decrease in inaccuracy, this can impede the development and training of models and require a vast number of system memories. Additionally, less data means that algorithms train faster [8–10].

In the present work, two different feature selection methods have been used, which are provided by the sci kit-learn Python library are ANOVA *F* test and Mutual Information. The feature selection is also applied to the wavelet decomposed signal and compares the classification accuracy with classification using raw data statistical features.

Table 1 Vibration data description for the fault classification

Case No.	Status	Class index/number of sample	Number of target class
Case 1	Presence of fault	Class = 0 (intact)/$n = 320$ Class = 1 (fault)/$n = 360$	2
Case 2	Type of fault	Class = 0(intact)/$n = 320$ Class = 1(inner race fault)/$n = 120$ Class = 2(ball fault)/$n = 120$ Class = 3(outer race fault)/$n = 120$	4
Case 3	Fault severity	Class = 0 (intact)/$n = 320$ Class = 1(0.007 inch fault size in inner race)/$n = 40$ Class = 2(0.014 inch fault size in inner race)/$n = 40$ Class = 3(0.021 inch fault size in inner race)/$n = 40$ Class = 4(0.007 inch fault size in ball)/$n = 40$ Class = 5(0.014 inch fault size in ball)/$n = 40$ Class = 6(0.021 inch fault size in ball)/$n = 40$ Class = 7(0.007 inch fault size in outer race)/$n = 40$ Class = 8(0.014 inch fault size in outer race)/$n = 40$ Class = 9(0.021 inch fault size in outer race)/$n = 40$	10

Wavelet Packet Transform

The wavelet packet transforms for any time-domain signal is acts as a computation process that involves approximation and assessment of details in signal on passing it through low and high pass filters. To find the temporary location of transient activities that happens during the observation of the development of a problem on the bearing's surface. Which is helpful for monitoring, and defect detection is two aspects of condition monitoring. The wavelet transform uses the time and scale window functions to characterize signals in the frequency–time domain Wavelet packets filter the incoming signal into ever more acceptable equal-width intervals; as a result, leading to sub-bands filtering. The frequency axis is separated into sub-bands [0, 1/2] at each level, j. At level j, the sub-bands in hertz are $\left[nF_s/2^{j+1}, (n+1)F_s/2^{j+1}\right]$, Where F_s is the sampling frequency. Compare to other wavelet techniques wavelet packets are superior at time–frequency analysis. Wavelet packets also have the advantage of having orthogonal transforms (when using an orthogonal wavelet). The following section shows that an orthogonal transform retains the signal's energy and distributes it among the coefficients [11, 12].

Feature Selection

ANOVA *F* Test

The analysis of variance (ANOVA) is a statistical tool to determine if the mean of more than one group differs significantly. The innovative aspect of the current work is the use of one-way ANOVA and the *F* test statistical test as a prediction method to evaluate harmony for feature selection and to describe the key characteristics to minimize the total data dimensionality of the feature space, with the main objectives being to decrease computational time complexity or to increase classification accuracy, or both [13]. When comparing more than one group of numerical data with only one independent and one dependent variable, "one-way ANOVA" is used. The goal is to see if the data are from distinct groups have the same mean. One-way traffic ANOVA presupposes that comprehensive data within the group have a normal distribution; however, it can also operate with somewhat skewed data from the norm. Either ANOVA examines the null hypothesis (H_0), which states that the means of all groups are equal, or it tests the null hypothesis (H_1), which states that at least one group's mean is different (H_1):

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots \mu_k \quad (1)$$

where $\mu_{(k)}$ represent the mean for groups, k represents the sum of all groups, H_0 and H_1 are standard hypothesis test symbols, with H_0 denotes the accepted hypothesis, and the

Fig. 1 A proposed methodology for rolling element bearing fault classification
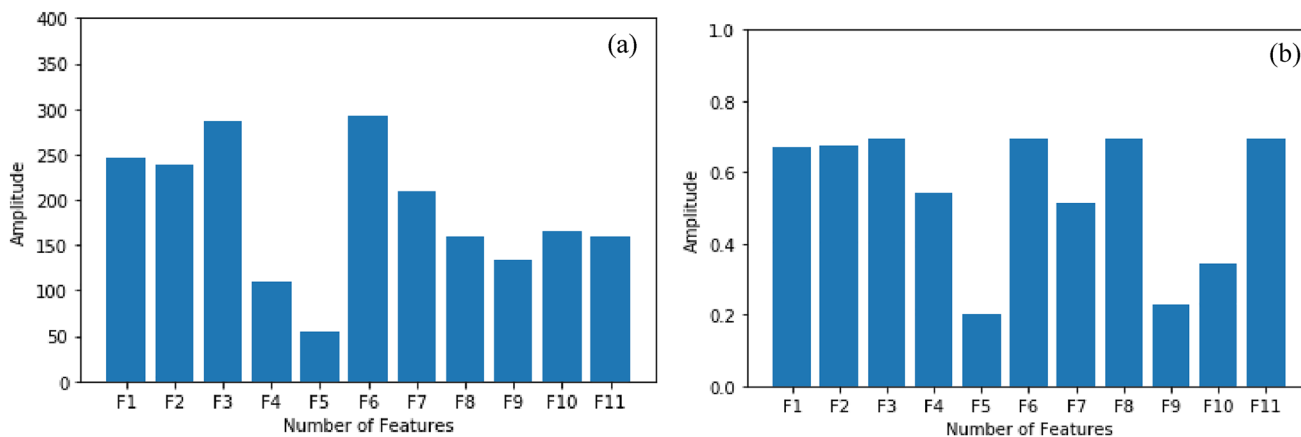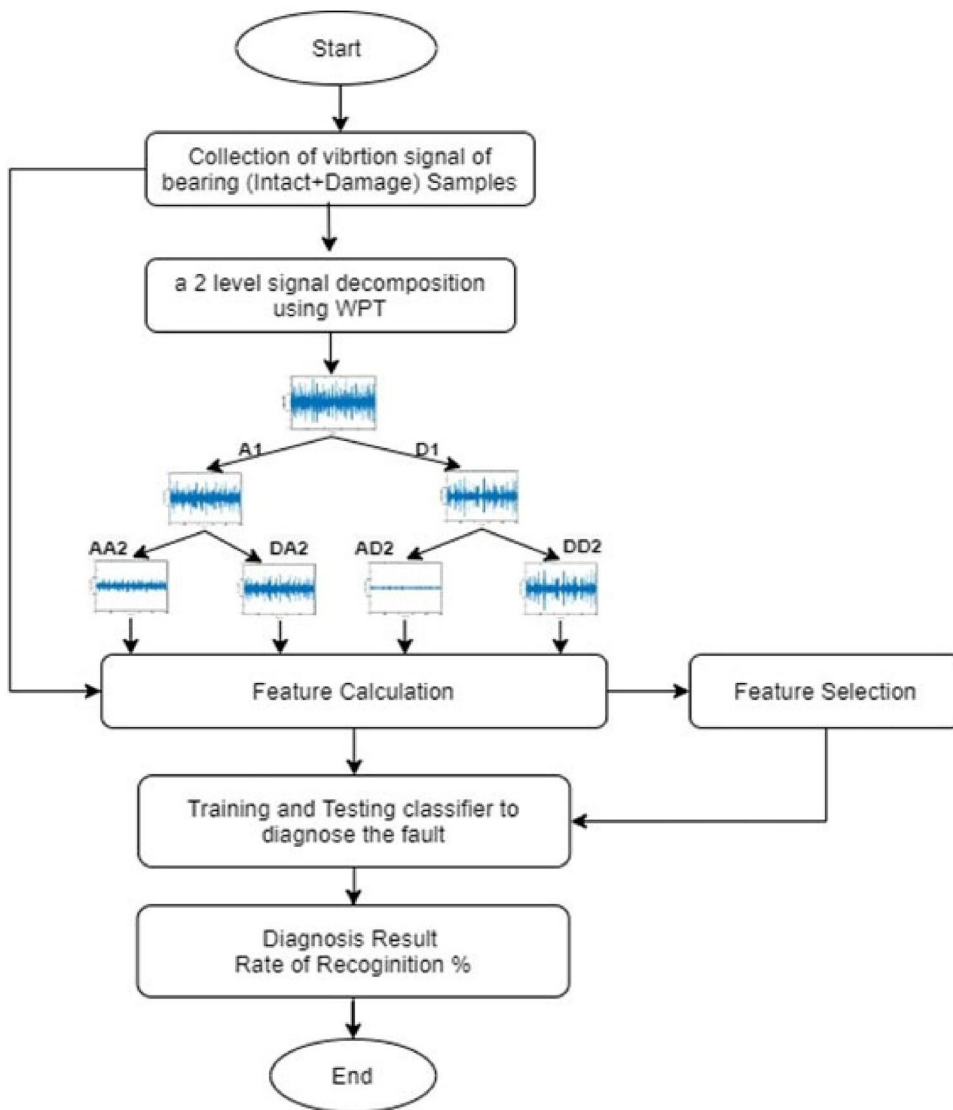
Fig. 2 Features score of the original data set **a** ANOVA *F* test **b** Mutual information

rejected hypothesis shows by H_1. In this approach, ANOVA divides the overall sum of squares (SST) into the sum of squares (SSR) because of the between-groups effect and the sum of squared errors (SSE):

$$SST = \sum \sum \left(y_{ij} - \bar{y}\right)^2 = \sum \sum \left(y_{ij} - \bar{y}_j\right)^2 + \sum n_j \left(\bar{y}_j - \bar{y}\right)^2, \quad (2)$$

where SSE and SSR are the first and second terms of the above equation, and $(\bar{y}_j)$ is the group means, (y_{ij}) is the ith data position within the group j, $\bar{y}$ is the total mean of the groups, n_j is the sample size of group j, put $j = 1,2,....,k$; and k is the total number of groups. The ratio of their variance between groups to variance within the group is assessed to see whether the groups' means are significantly distinct from one another. Although a larger ratio clearly distinguishes the groups, ANOVA based on test statistics with an F distribution and $(N - k, k - 1)$ degrees of freedom (DOF) to make calculations easier:

$$F = \frac{SSR/(k - 1)}{SSE/(N - k)} = \frac{MSR}{MSE} \sim F_k, N - k. \quad (3)$$

MSE stands for mean squared error, MSR is for mean squared treatment, N shows the total number of observations, and k represents the number of groups. Finally, the p value is determined using the CDF of the F distribution. The rejection of the null hypothesis occurs when the result of p is less than the significance level, indicating that there must be at least one group with a different mean.
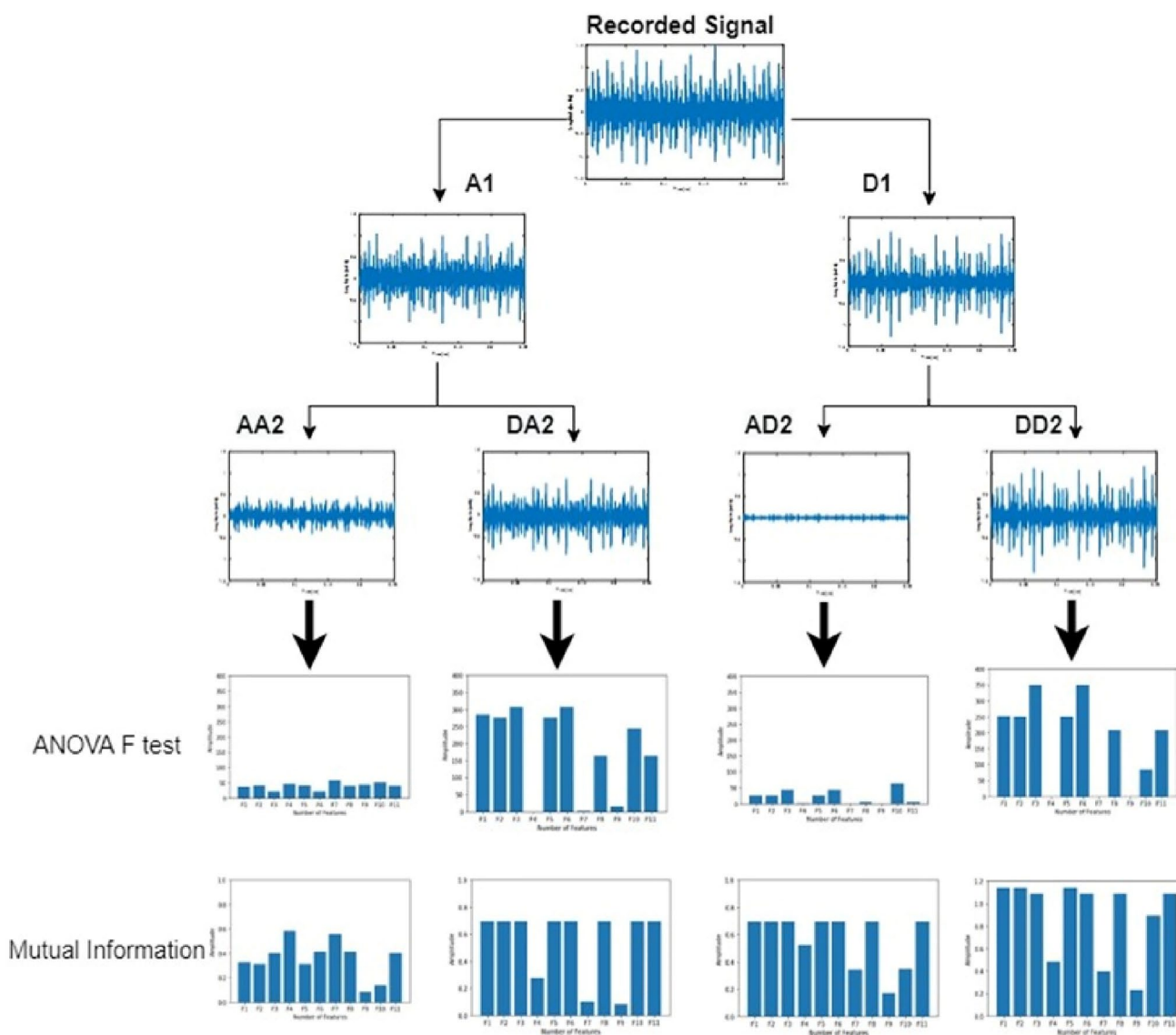

Fig. 3 Features score of the WPT decomposed data set ANOVA F test and mutual information

Mutual Information

The mutual information technique (MIT) is crucial for feature selection in ball-bearing fault diagnosis. MIT used for high-order and nonlinear transformation statistics extraction. As a result, we are considering MIT for feature selection to equalize the length of the samples. Similarly, we can effectively reduce dimensions by selecting features using the nonlinear relationship in multi-dimensional feature space. The probability density estimation technique substantially influences MI calculating, implying whether the method can efficiently and effectively improve feature selection accuracy to express typical features. As a result, in fault diagnosis using MIT feature selection, probability density estimation is an appropriate critical method. We first compute the MIT value using a probability density estimation method to extract the relationship between linear and nonlinear variables through MI matrices [14, 15].

The mutual information MI ($x;y$) is a quantity between two discrete random variables x and y that finds the two variables' mutual dependence and can be calculated as:

$$\mathrm{MI}(X_v; Y_v) = \iint S(X_v, Y_v) \log \frac{S(X_v, Y_v)}{S(X_v), S(Y_v)} \mathrm{d}X_v \mathrm{d}y_v, \quad (4)$$

where $S(X_v, Y_v)$ is the joint probability density function of X_v and Y_v. $S(X_v)$ and $S(Y_v)$ are the marginal probability density functions of X_v and Y_v, respectively. The estimation of joint probabilities using predictors depends on kernel, density binning, or nearest neighbors when at least one continuous random variable is present. The nearest neighbor's estimator outperforms the other two methods because it is an adaptive estimator and data-efficient. As a result, the present study

| Data	All feature	Training/testing samples	Accuracy (%) LR	SVC
Original	11	455/225	94.67	93.33
AA2	11	455/225	71.11	68.00
AD2	**11**	**455/225**	**100**	**100**
DA2	11	455/225	83.11	81.77
DD2	**11**	**455/225**	**100**	**100**

employs the nearest neighbor's estimator to estimate MIT on process parameters.

The goal of parameter selection based on MIT in classification is to determine a set R of n parameters X_n that has an enormous dependency on class T.

$$\max d(R, T), d = \mathrm{MI}\big(\{X_n|, n = 1, \ldots \ldots, m\}; T\big) \quad (5)$$

The criteria for max-dependency (max$d$) are round about using simplified rules such as the mutual information criteria because the functions for joint probability density are complex to estimate in practice due to a lack of samples. The univariate mutual information MI ($X_n; T$) between a variable X and the class T is the most straightforward criterion. The greater the value of MI ($X_n; T$), the more important X_n. In the classification it can be written as:

$$\mathrm{MI}(X_n; T) = P(X_n, T) \log \frac{P(X_n, T)}{p(X_n), P(T)} \mathrm{d}x_n \mathrm{dt}. \quad (6)$$
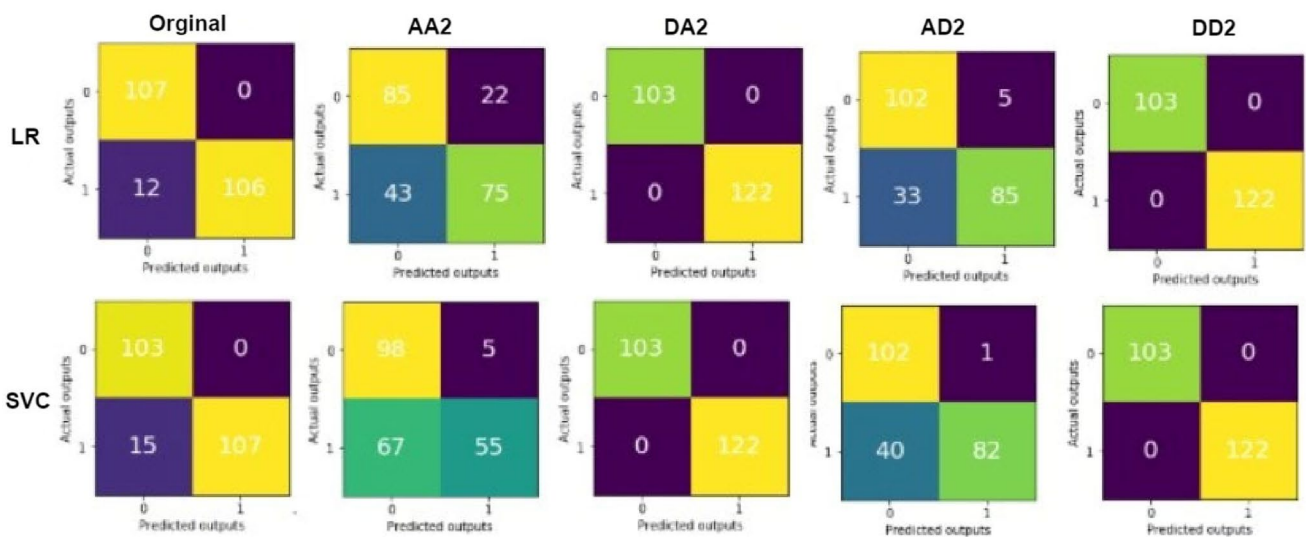
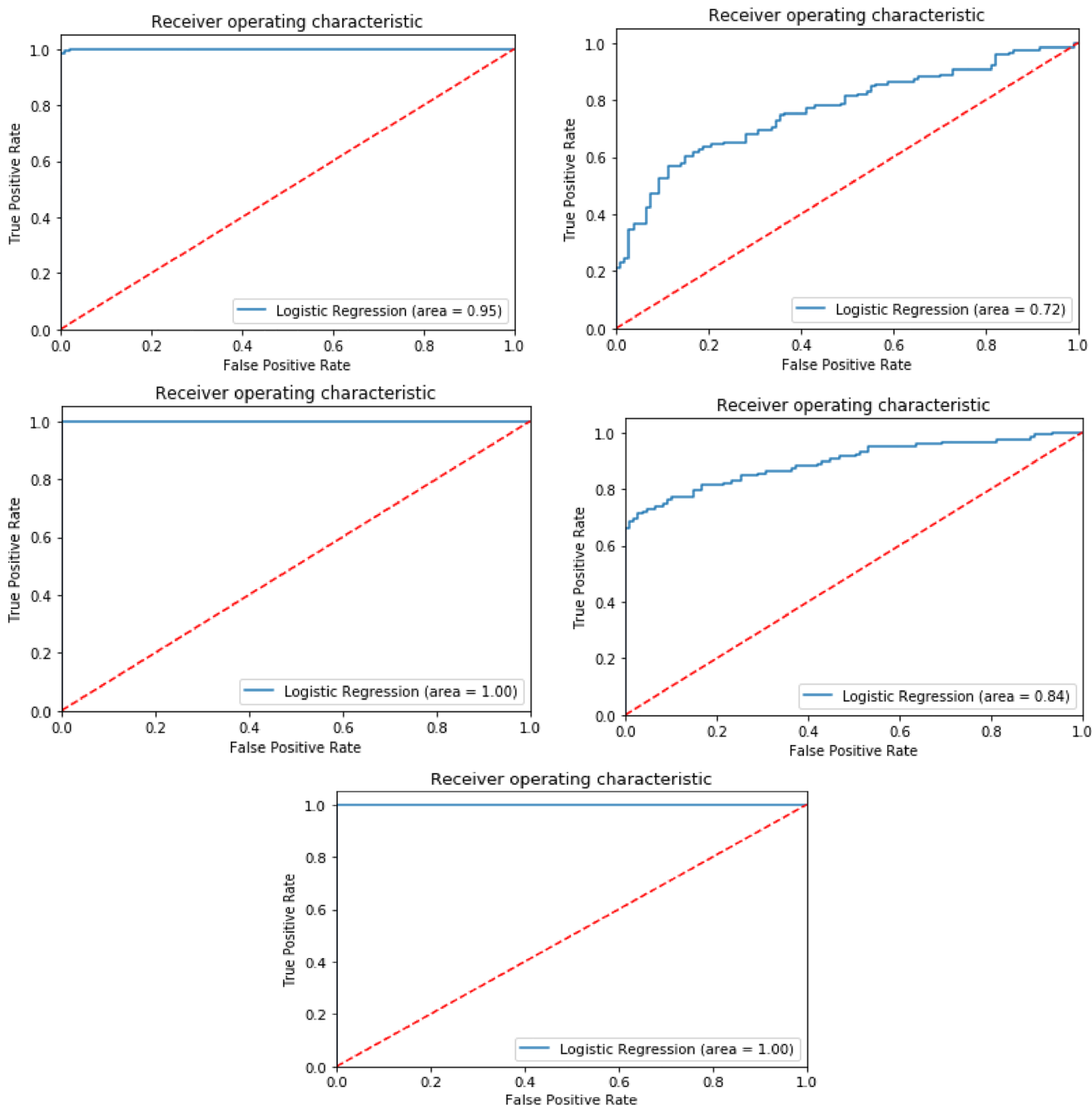**Fig. 4** Fault isolation results using all the features for case 1

Fig. 5 ROC curve for the original and decomposed signals

All the possible values were represented by X_n and T along with any values of X_n and T. The extension of the nearest neighbors, using estimator probability functions, is computed between a continuous and a discrete function. The probability diffusion functions are calculated using a continuous and discrete elements extension of the nearest neighbor's predictors.

Classification Techniques

Logistic Regression (LR)

A supervised machine learning algorithm, logistic regression, is used for a collection of features (or inputs), X, the target variable (or output), Y, that can only accept discrete values in a classification problem. When the dependent variable is nonparametric, logistic regression is a version of

Table 3 Classification accuracy for LR and SVC with selected features for case 1

Data set	Optimal number of feature	Accuracy (%) LR	Optimal number of feature	Accuracy (%) SVC
Original	10	96.30	5	97.20
AA2	11	71.80	8	67.20
AD2	**6**	**100**	**6**	**100**
DA2	11	85.60	3	83.70
DD2	**4**	**100**	**3**	**100**

ordinary regression (represented by the occurrence or non-occurrence of some output events, usually coded between 0 and 1) [16]. The purpose of logistic regression is to identify the best-fitting model to represent the connection between a set of independent factors and a dichotomous characteristic of the dependent variable [14]. The dependent variable in the logistic regression approach is the chance of an event occurring; thus, the output has a discrete range of respondents confined between 0 and 1. The logistic function is described as follows:

$$\text{prob(event)} = p(\bar{x}) = \frac{1}{1 + e^{-g(\bar{x})}} = \frac{e^{g(\bar{x})}}{1 + e^{g(\bar{x})}}, \quad (7)$$

where $p(x)$ p is some output vent probability, $\vec{x}(x_1, x_2, x_3, \ldots, x_k)$ represents the input vector corresponding to the predictors (independent variables), and $g(x)$ represents the logit model. Multiple logistic regressions' logit model can be stated as:

$$g(x_1, x_2, \ldots, x_k) = a + b_1 x_1 + b_2 x_2, \ldots b_k x_k, \quad (8)$$

where $g(x)$ is a linear combination of the independent variables $X_1; X_2; \ldots X_k$ and $a; b_1; b_2; \ldots b_k$ are known as the regression coefficient. Logistic regression employs maximum likelihood estimation after transforming the dependent into a logit variable to determine the parameter $a; b_1; b_2; \ldots b_k$, after converting the dependent into a logit variable. The probability of failure for run-to-failure bearing data is estimated using logistic regression in this research. This failure probability depicts failure progression from incipient failure (encoded as 0) to complete failure circumstances (denoted as 1).

Support Vector Classifier (SVC)

In a large or indefinite dimensional space, a support vector machine creates a hyper-plane or set of hyper-planes that can be used for classifications, regression, and other tasks. Instinctively, the hyper-plane with the most significant distance to the adjacent training data points of any class (so-called operating margin) achieves a substantial separation because the more extensive the margin, the lesser the generalization error of the classifier [13, 17].

SVC solves the following problem:

$$\min_{\omega, b, \zeta} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^{n} \zeta_i. \quad (9)$$

Subjected to $y_i(\omega^T \phi(x_i) + b) \geq 1 - \zeta_i,$ $\zeta_i \geq 0, i = 1, \ldots, n.$

The purpose is to expand the margin (by minimizing $(\|\omega\|^2 = \omega^T \omega)$) while incurring a penalty when a sample is misclassified or within the margin boundary. Ideally, the value $y_i(\omega^T \phi(x_i) + b)$ would be ≥ 1 for every sample and it denotes perfect anticipation. However, because issues are rarely entirely separable with a hyper plane, we allow some samples to be separated from their correct margin boundary by a distance ξ_i. As a result, the penalty term C works as an inverse regularization parameter, controlling the severity of the penalties.

Proposed Methodology

In the present paper, the induction motor bearing fault classification has been proposed in three categories, i.e., presence of faults, type of faults, and the fault severity, as mentioned in Table 1. To obtain the meaningful signal, a filtering procedure is required. A two-level WPT has been used to extract the signal, which has been divided into four sub-bands with frequency ranges of 0–1500 Hz, 1500–3000 Hz, 3000–4500 Hz, and 4500–6000 Hz, respectively, and designated as AA2, DA2, AD2, and DD2. For four different load situations, the signal decomposition has been carried out for intact, three different damaged conditions, and three different fault sizes.

The process of finding and choosing a subgroup of input features that are most appropriate for the target variable is known as feature selection. Feature selection is often simple when using real-valued input and output data, such as Pearson's correlation coefficient. Still, it can be demanding when operating with a numerical input variable and a categorical target variable. Efficient diagnosis and prognosis can be achieved by selecting the most essential and sensitive features. Incorrect and inaccurate features degrade the overall reliability of fault diagnosis and prognosis approaches, making it impossible to anticipate actual bearing conditions.

The statistical feature calculation in the current work makes use of the unprocessed and sub-band data obtained from WPT. Each recorded vibration signal has had eleven statistical features derived from it. This results in a feature set of 680*11 for unprocessed and each sub-band data,
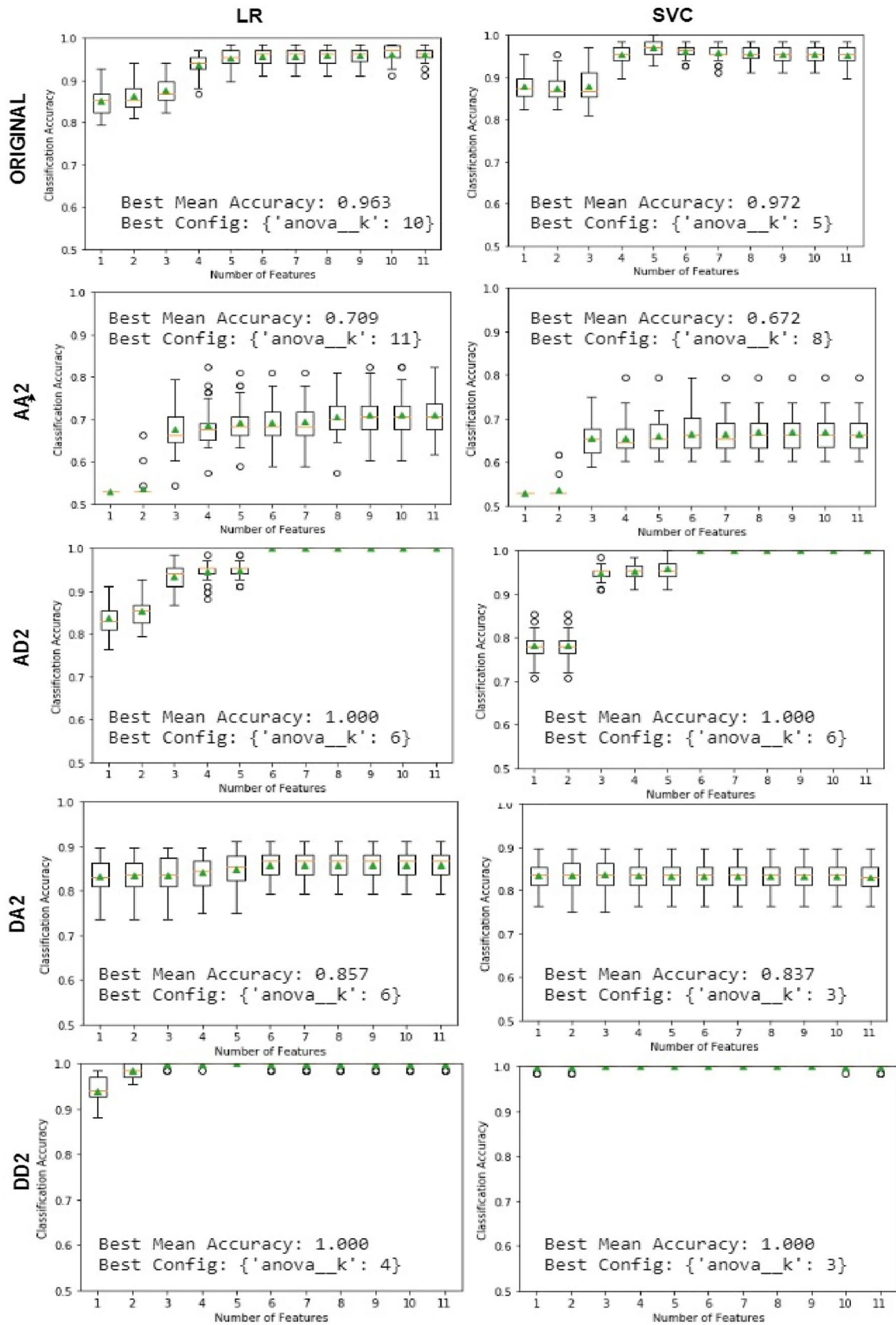
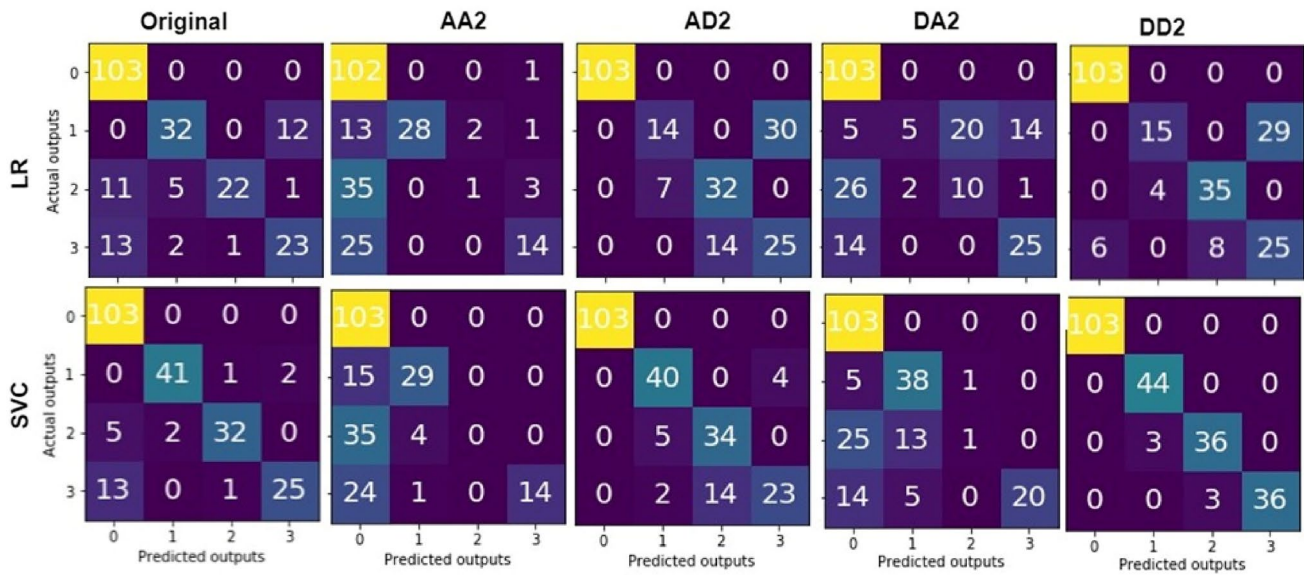**Fig. 6** Grid search for the optimum number of features for case 1

Fig. 7 Fault isolation results using all the features for case 2

Table 4 Classification accuracy using all the features as input for case 2

Data set	With all feature	Training/testing samples	Accuracy (%)	
			LR	SVC
Original	11	455/225	77.33	87.5
AA2	11	455/225	62.66	62.2
AD2	11	455/225	**81.33**	87.55
DA2	11	455/225	70.22	72.80
DD2	11	455/225	76.00	**96.4**

which is ready for classification. For the purpose of gathering additional data and correctly classifying problems, features are connected for their applicability and reactivity to various defects. When a categorical target variable is present, the ANOVA F test and mutual information statistics are the two most often utilised feature selection techniques for numerical input data. Logistic Regression and SVC training and testing make use of the features selected using the feature selection methods. The methodology for the suggested work is shown in Fig. 1.

Result and Discussions

The three scenarios were taken into consideration in this study to evaluate the induction motor malfunction status using vibration signals. Case 1 examines the presence of a fault, i.e., intact and fault condition of bearing. As per the proposed methodology initially, from the 680 original data sets, 11 features have been calculated from each data set, of

which 320 datasets are for the intact bearing and 360 datasets are for the damaged bearing. All eleven input variables are numerical types. The target value is 0 and 1, respectively, for intact and fault conditions. The prepared data have been applied for the ANOVA F test and mutual information to investigate the optimal number of features. A bar chart of the feature importance scores for each input feature is created and shown in Fig. 2. Figure 2 shows the scores of the ANOVA F test for each variable (more prominent is superior) and plots the scores for each variable as a bar graph to get a scheme of how many features we should select. The results of this test can be owned for feature selection, where those features that are independent of the target variable can be detached from the dataset. From Fig. 2, we can conclude that features F4 and F5 are irrelevant as their scores are low. In this case, it has been observed that some features stand out as perhaps being more relevant than others, with much larger test statistic values. The features F3 and F6 might be the most suitable (according to the test), and perhaps six of the eleven input features are the more relevant from the ANOVA F test. F5 and F9, and F10 have minor importance from the mutual information technique feature due to low scores.

WPT decomposes each original set into four sub-bands, as shown in Fig. 3. Wavelet packets filter the incoming signal into progressively finer equal-width intervals, resulting in sub-band filtering. WPT is a helpful method for detecting and discriminating transient elements with high-frequency characteristics because of the sub-bands. The eleven features were calculated again from four sub-band signals separately and searched for significant features using the ANOVA F test and Mutual Information.
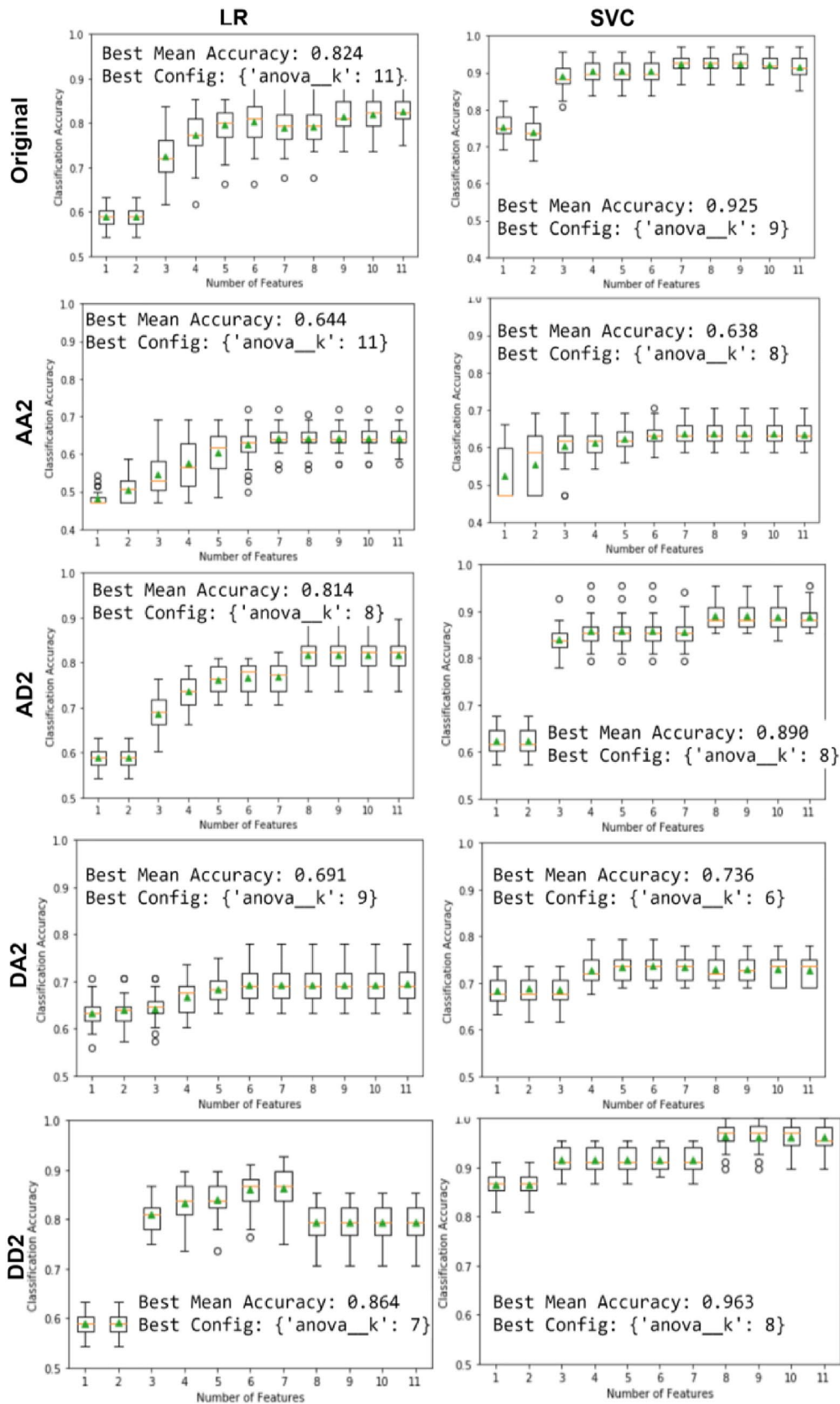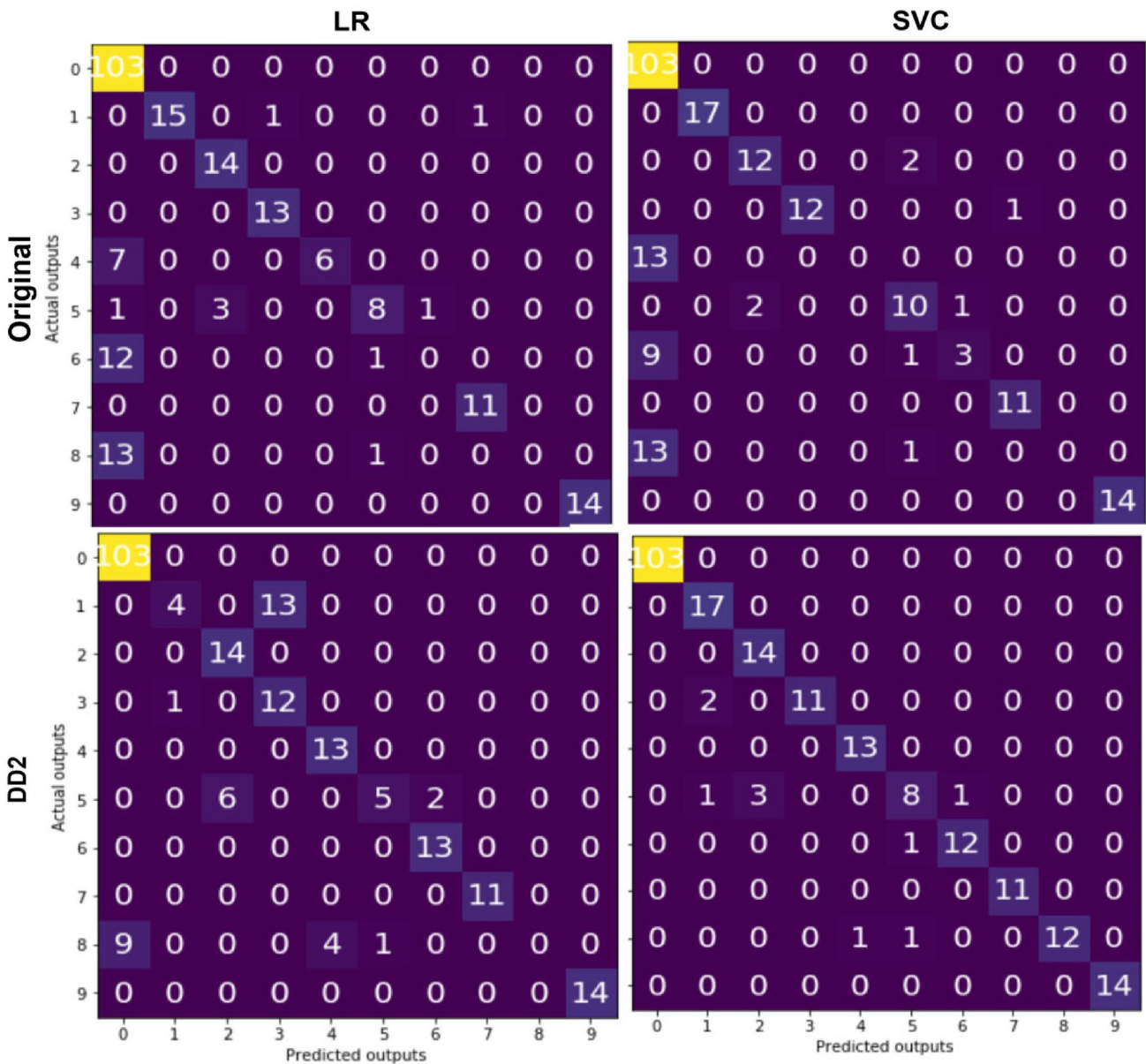
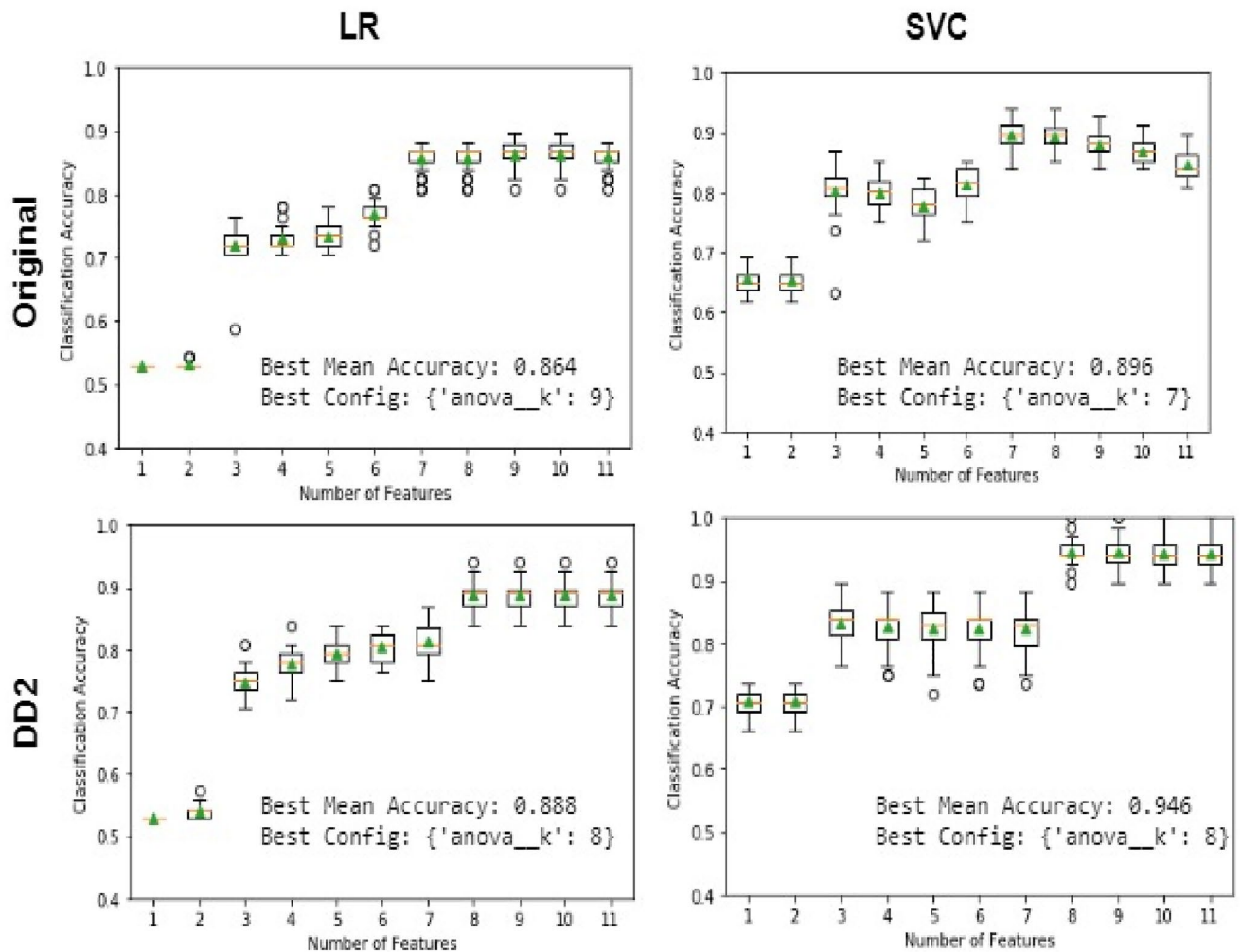**Fig. 8** Grid search for the optimum number of features for case 2

Table 5 Classification accuracy using the selected feature as input for case 2

Data set	Optimal number of feature	Accuracy (%) LR	Optimal number of feature	Accuracy (%) SVC
Original	11	82.40	9	92.50
AA2	11	64.40	8	63.80
AD2	8	81.40	8	89.00
DA2	9	69.10	6	73.60
DD2	**7**	**86.40**	**8**	**96.30**

From Fig. 3, we observed that in ANOVA F test, signal AA2 and AD2 score extremely low, and all the others features show approximately the same value. In signal DA2 and DD2, the features F4, F7, and F9 are insignificant. Feature selection by mutual information found that the feature score from DD2 high compares to AA2, DA2, and AD2. Feature F9 F10 in the signal AA2 and F4, F7, and F9 found less score than the other features in the signal DA2, AD2, and DD2. The feature score has determined that some features are irrelevant or useless because they have a low F score or have a small impact on classification accuracy. It has also been observed that the ANOVA F test gives a better selection of features compared to mutual information.

Fig. 9 Fault isolation results using all the features for case 3

Fig. 10 Grid search for the optimum number of features for case 3

The performance of feature selection on numerical input data for a classification predictive modelling challenge must now be investigated. Using the chosen features, we created a model, and then we compared the results. In this section, a logistic regression (LR) and support vector classification (SVC) model with all features are evaluated and collated to a model built from features selected by the ANOVA *F* test and those features selected via mutual information. Logistic regression is a good technique for feature selection as it can perform better if irrelevant features are removed from the model.

A total of 680 data samples are prepared for the training and testing phases to perform a two-class classification problem that covers both intact and defect-bearing circumstances. The flow chart in Fig. 1 depicts the methods provided for diagnosing and classifying various faults in this context. Figure 4 shows the confusion matrix of the original signal and the decomposed signal obtained from logistic regression and support vector classifier. Out of 680

data set, 455 data set has been used for training, and 225 have been used for testing purpose. Table 2 shows the percentage accuracy of classifier LR and SVC when all the 11 features have been considered for all five signals. From the results, it has been concluded that the signal AD2 and DD2 give 100% accuracy for both the classifier which is the best among all the signal for both the classifier LR and SVC. The original signal accuracy is 94.67% for LR and 93.33% for SVC under the same condition. The signal AA2 gives lower accuracy 71.11% and 68%% for LR and SVC, respectively. Similarly, the classification accuracy obtained from signal DA2 is 83.11% and 81.77% for LR and SVC classifiers, respectively. From the obtained results, it may be concluded that when all 11 features were taken into consideration AD2 and DD2 gave high 100% accuracy from both the classifier; however, for the rest signal, LR performs better than the SVC for binary classification.

The best curve obtained from the original, AD2, and DD2 signals equates to 100% accuracy, as shown in Fig. 5

Table 6 Comparison table demonstrating the implication of present study with reported literature

References	Number of features	Feature selection techniques	Classifier used	Peak efficiency	Optimal number of features	Class of problem
Kavathekar et al. [18]	10	NIL	Seven classifier	75%	10	4
Wu et al. [19]	80	Fisher score and Mahalanobis	SVM	98.5%	10	6
Wang et al. [20]	100	Statistically locally linear embedding	CART, KNN, and SVM	97.26	22	NIL
Li et al. [21]	180	Laplace score	SVM-BT	100%	NIL	12
Zhang et al. [22]	20	PCA, LPP, LDA, and ISOP	Neighborhood preserving embedding and SOM	99%	4	10
Pablo et al. [23]	35	Relief and information gain	SVM and random forest	98.38%	10	NIL
Dubey et al. [24]	4	NIL	Neural network	100%	NIL	3
Vakharia et al. [25]	10	NIL	SVM	97.50%	NIL	4
Sharma et al. [26]	10	NIL	SVM and ANN	78	7	7
Van et al. [27]	78	PSO	KNN, PNN, or SVM	98.58	20	7
Deng et al. [28]	160	NIL	EWTFSFD and EMDFSFD	70	14	4
Babouri et al. [29]	60	PCA	MSPCA-FL	98.50	9	10
Present work	**11**	**ANOVA and mutual Information with WPT**	**SVC**	**100%**	**3**	**2**
			LR	**100%**	**4**	
			SVC	**96.3%**	**8**	**4**
			LR	**86.4%**	**7**	
			SVC	**94.6%**	**8**	**10**
			LR	**88%**	**8**	

ROC curve for the original and decomposed signal. This section explores improvement in the classifier's performance using the grid search approach to reduce the number of features and achieve the same or higher accuracy when all features are used for classification. The first step is to define a series of modeling pipelines to evaluate. Each channel describes data preparation techniques and ends with a model that takes the transformed data as input. To determine which features produce the best performing model, a variety of various numbers of selected features have been carefully tested. In a grid search, the k argument to the SelectKBest constructor tells the selector that it must score the variables according to an F score calculated starting from Pearson's correlation coefficient between each feature and the target variable. Following the feature selection, a LR and SVC will be run on the chosen features. Then executes a grid search on the quantity of Python features. Using repeated stratified k-fold cross-validation to assess model configurations on classification problems is a useful practice. In this study, three repeats of tenfold cross-validation have been used for all three cases. For each cross-validation fold, we may describe a pipeline that correctly organised the feature selection to change the training set and applied it to the train set and test set.

The evaluation grid can, therefore, be defined as a range of values from 1 to 11. The classification accuracy of both classifiers is shown in Table 3.

The ANOVA F test is used to run grid searches with various features that have been chosen, and each modelling pipeline is assessed using repeated cross-validation. The grid search technique for the most features that provides the best accuracy is shown in Fig. 6. The best features for five signals and the classification accuracy for the classifiers LR and SVC were achieved. In the case of LR, the 100% accuracy was obtained with 4 number features, and in SVC, only 3 number features from the DD2 signal for 100% accuracy. The same 100% accuracy was obtained from the signal AD2 with six features for both the classifiers. The best accuracy is achieved from the original signal at 96.3% with ten features in LR and 97.2% with five features from SVC. From the obtained classification accuracy, it can be concluded that the SVC classifier gives the best performance compared to the LR classifier.

For case 2, the various fault categories have been taken into account when classifying. It consists of four unique classes: the inner race fault (IRF), ball fault (BF), and outside race fault (ORF), as well as one that is in its intact state. For intact, IRF, BF, and ORF, the target values are 0, 1, 2, and 3. Figure 7 shows the multiclass classification task that covers four individual rolling conditions. Table 4 shows that the LR classifier obtains an accuracy of 81.33% for the WPT signal AD2 and 96.4% for the WPT signal DD2 when using SVC when considering all eleven attributes. SVC classifier performs better when the WPT filtered signal is used to determine the type of bearing defect.

The most precise fault type detection is achieved using a grid search approach to determine the ideal number of features. The grid search outcomes for the LR and SVC models with the ideal number of features are shown in Fig. 8. Table 5 provides a summary of the findings. According to the results obtained to categories the faults in the original signal, the optimal configuration is made up of nine features with the best accuracy of 92.5% when using SVC and 11 features with an accuracy of 82.4% when using LR. However, in the case of WPT, the DD2 signal SVC model with eight features and 96.3% accuracy provided the optimum structure. With a maximum of 7 features, LR provides an accuracy of 86.4% for the identical signal.

It is evident from Case 1 and Case 2 that the WPT signal DD2 yielded the best accuracy and the ideal amount of characteristics. In case 3, the WPT signal DD2 has been taken into consideration for additional processing to determine the fault severity. Figure 9 shows the confusion matrix for case 3 for LR and SVC classifier. The original signal gives 81.77% and 80.88% accuracy from LR and SVC classifier, respectively, when all features are considered. The obtained accuracy for signal DD2 is 84% and 95.55% from LR and SVC, respectively, when all features consider as input to both the classifier. Figure 10 shows the optimum features to detect the successful fault severity level. The best accuracy achieves 94.6% with an optimum number of features 8 for the DD2 signal using the SVC classifier. (Table 6) compares current work to previous work in the litirature of condition monitoring of rotating machines. According to the table in the current work, the better accuracy with the fewest features has been obtained.

Conclusions

In this study, the vibration signals from the induction motor bearing were used to classify the three cases of fault presence, fault kind, and fault severity. Eleven statistical features were computed from the original signal and the two-level wavelet decomposed signal for Case 1, Case 2, and Case 3 to classify the data. The best number of features for classification is selected using mutual information and the ANOVA F test from the sub-band having high F score. Further, a Logistic Regression and Support Vector Classifier were used to classify each of the three cases, and the results were compared for each case using both the full set of features and the chosen features. In comparison to LR classifier, sub-band signal DD2 with SVC gives the best results for all three cases from the results obtained with the full set feature. The grid search method with SVC produced the greatest results, with three features providing 100% accuracy for case 1; eight features providing 96.3% accuracy for case 2; and eight features providing 94.6% accuracy for case 3. Therefore, it can be concluded that the suggested methodology can be used in practical to detect bearing faults in induction motors while obtaining the ideal number of features and greater accuracy.

References

1. Tavner P, Ran L, Penman J, Sedding H (2008) Condition monitoring of rotating electrical machines, The Institution of Engineering and Technology, London, pp 1–306
2. de Almeida LF, Bizarria JW, Bizarria FC, Mathias MH (2014) Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron. J Vib Control. https://doi.org/10.1177/1077546314524260
3. Samanta B, Al-Balushi KR, Al-Araimi SA (2003) Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Eng Appl Artif Intell 16(7–8):657–665. https://doi.org/10.1016/j.engappai.2003.09.006
4. Kankar PK, Sharma SC, Harsha SP (2011) Fault diagnosis of ball bearings using continuous wavelet transform. Appl Soft Comput 11(2):2300–2312. https://doi.org/10.1016/j.asoc.2010.08.011
5. Vakharia V, Gupta VK, Kankar PK (2015) A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings. JVC/J Vib Control 21(16):3123–3131. https://doi.org/10.1177/1077546314520830
6. "Bearing Data Center." https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website. Accessed 15 Oct 2019
7. Patel RK, Giri VK (2017) ANN based performance evaluation of BDI for condition monitoring of induction motor bearings. J Inst Eng Ser B 98(3):267–274. https://doi.org/10.1007/s40031-016-0251-7
8. Vakharia V, Gupta VK, Kankar PK (2017) Efficient fault diagnosis of ball bearing using ReliefF and random forest classifier. J Brazilian Soc Mech Sci Eng. https://doi.org/10.1007/s40430-017-0717-9
9. Kumar HS, Pai SP, Sriram NS, Vijay GS (2016) Rolling element bearing fault diagnostics: Development of health index. Proc Inst Mech Eng Part C J Mech Eng Sci. https://doi.org/10.1177/0954406216656214
10. Gangsar P, Tiwari R (2020) Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: a state-of-the-art review. Mech Syst Signal Process 144:106908. https://doi.org/10.1016/j.ymssp.2020.106908
11. Anwarsha A, Narendiranath Babu T (2022) A review on the role of tunable Q-factor wavelet transform in fault diagnosis of rolling element bearings. J Vib Eng Technol 10:1793–1808. https://doi.org/10.1007/s42417-022-00484-1
12. Li X, Jia L, Yang X (2015) Fault diagnosis of train axle box bearing based on multifeature parameters. Discret Dyn Nat Soc. https://doi.org/10.1155/2015/846918
13. Elssied NOF, Ibrahim O, Osman AH (2014) A novel feature selection based on one-way ANOVA F-test for e-mail spam classification. Res J Appl Sci Eng Technol 7(3):625–638. https://doi.org/10.19026/rjaset.7.299
14. Li B, Zhang PL, Tian H, Mi SS, Liu DS, Ren GQ (2011) A new feature extraction and selection scheme for hybrid fault diagnosis

of gearbox. Expert Syst Appl 38(8):10000–10009. https://doi.org/10.1016/j.eswa.2011.02.008

15. Battiti R (1994) Using mutual information for selecting features in supervised neural net learning. IEEE Trans Neural Netw 5(4):537–550. https://doi.org/10.1109/72.298224

16. Pandya DH, Upadhyay SH, Harsha SP (2013) Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. Soft Comput 18(2):255–266. https://doi.org/10.1007/s00500-013-1055-1

17. Shahriar M, Ahsan T, Chong U (2013) Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis. EURASIP J Image Video Process 2013(1):29. https://doi.org/10.1186/1687-5281-2013-29

18. Kavathekar S, Upadhyay N, Kankar PK (2016) Fault classification of ball bearing by rotation forest technique. Procedia Technol 23:187–192. https://doi.org/10.1016/j.protcy.2016.03.016

19. De Wu S, Wu CW, Wu TY, Wang CC (2013) Multi-scale analysis based ball bearing defect diagnostics using mahalanobis distance and support vector machine. Entropy 15(2):416–433. https://doi.org/10.3390/e15020416

20. Wang X, Zheng Y, Zhao Z, Wang J (2015) Bearing fault diagnosis based on statistical locally linear embedding. Sensors (Switzerland) 15(7):16225–16247. https://doi.org/10.3390/s150716225

21. Li Y, Xu M, Wei Y, Huang W (2016) A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. Meas J Int Meas Confed 77:80–94. https://doi.org/10.1016/j.measurement.2015.08.034

22. Zhang S, Li W (2014) Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM. Math Probl Eng. https://doi.org/10.1155/2014/781583

23. Sánchez RV, Lucero P, Vásquez RE, Cerrada M, Macancela JC, Cabrera D (2018) Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN. J Intell Fuzzy Syst 34(6):3463–3473. https://doi.org/10.3233/JIFS-169526

24. Dubey R, Agrawal D (2015) Bearing fault classification using ANN-based Hilbert footprint analysis. IET Sci Meas Technol 9(8):1016–1022. https://doi.org/10.1049/iet-smt.2015.0026

25. Vakharia V, Gupta VK, Kankar PK (2014) A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings. J Vib Control. https://doi.org/10.1177/1077546314520830

26. Sharma A, Amarnath M, Kankar PK (2016) Feature extraction and fault severity classification in ball bearings. JVC/J Vib Control 22(1):176–192. https://doi.org/10.1177/1077546314528021

27. Van M, Kang HJ (2015) Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection. IET Sci Meas Technol 9(6):671–680. https://doi.org/10.1049/iet-smt.2014.0228

28. Deng W, Zhang S, Zhao H, Yang X (2018) A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing. IEEE Access 6:35042–35056. https://doi.org/10.1109/ACCESS.2018.2834540

29. Babouri MK, Djebala A, Ouelaa N, Oudjani B, Younes R (2020) Rolling bearing faults severity classification using a combined approach based on multi-scales principal component analysis and fuzzy technique. Int J Adv Manuf Technol 107(9–10):4301–4316. https://doi.org/10.1007/s00170-020-05342-6