Engineering Science and Technology, an International Journal 18 (2015) 59–69

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

journal homepage: http://www.elsevier.com/locate/jestch

Full length article

Determination of minimum sample size for fault diagnosis of automobile hydraulic brake system using power analysis

V. Indira [a], R. Vasanthakumari [b], R. Jegadeeshwaran [c,*], V. Sugumaran [c]

[a] Department of Mathematics, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry, India
[b] Department of Mathematics, Villianur College for Women, Villianur, Puducherry, India
[c] School of Mechanical and Building Sciences, VIT University Chennai Campus, Chennai 600127, India

ARTICLE INFO

Article history:
Received 23 May 2014
Received in revised form
5 September 2014
Accepted 23 September 2014
Available online 4 November 2014

Keywords:
Fault diagnosis
Machine learning
Power analysis
Vibration signals
Minimum sample size
Statistical features

ABSTRACT

Hydraulic brake in automobile engineering is considered to be one of the important components. Condition monitoring and fault diagnosis of such a component is very essential for safety of passengers, vehicles and to minimize the unexpected maintenance time. Vibration based machine learning approach for condition monitoring of hydraulic brake system is gaining momentum. Training and testing the classifier are two important activities in the process of feature classification. This study proposes a systematic statistical method called power analysis to find the minimum number of samples required to train the classifier with statistical stability so as to get good classification accuracy. Descriptive statistical features have been used and the more contributing features have been selected by using C4.5 decision tree algorithm. The results of power analysis have also been verified using a decision tree algorithm namely, C4.5.

1. Introduction

The brake system is an essential component in an automobile to promote the highest degree of safety for the persons inside the vehicle and others moving on the road. Brake failure is crucial not only for the driver and passengers but also for automobile manufacturers. Fault diagnosis is an important process in preventive maintenance of hydraulic brakes. It avoids serious damage if defects occur to the component during operation. Prevention is better than cure. Early detection of the defects, therefore, is crucial to prevent the system from malfunction that could cause damage to entire system or accident. A fault diagnosis model can predict the condition of the system at any time and it avoids unexpected failures.

In this paper, only vibration signals of good and nine faulty conditions of a hydraulic brake system were considered for fault diagnosis. The characterization of the vibration signals was achieved by machine learning approach. The important two activities in machine learning approach are training and testing the classifier. To model the fault diagnosis problem as a machine learning problem, a large number of vibration signals are required for each condition considered for study. It may be possible to acquire any number of vibration signals for good condition; however, it is very difficult to acquire signals of different faulty conditions. Actually, the signals are to be taken from the system where the fault occurred naturally during operation. The difficulties involved in carrying out this process forces the fault diagnosis engineer to make a compromise. Taking many vibration signals from one specimen having a typical intended fault is one level of compromise in practice. Another level of compromise is that taking vibration signals from the system, where the required type of fault is simulated onto it. To overcome these problems, one should know the number of samples to be used for training to get good classification accuracy. Also, the signals can be acquired from the system where the fault has occurred naturally, only if one knows the minimum number of samples required to train the classifier to get good classification accuracy. Indeed, any results obtained out of these signals would be more practical and realistic in nature. Hence knowledge about the optimum number of samples required for building a model or training a classifier is very essential. In such situations, a study on determination of minimum sample size is highly desirable. Characteristics of the sample have a direct effect on power. Highly diverse samples will require adjustments in sample size. Adequate power is hard to achieve when results must

* Corresponding author. Tel.: +91 44 39931335; fax: +91 44 39932555.
E-mail addresses: indira_manik@yahoo.co.in (V. Indira), vasunthara1@gmail.com (R. Vasanthakumari), krjegadeeshwaran@gmail.com (R. Jegadeeshwaran), v_sugu@yahoo.com (V. Sugumaran).

Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2014.09.007

be very accurate. Very high confidence levels require very large samples. A study with insufficient power may lead the researcher to abandon potentially useful samples. Power analysis is the best method to avoid these serious errors.

In machine learning, a model built with large sample size would be robust. During implementation, it becomes necessary to know the number of samples required to build a classifier with statistical stability. Many researchers have taken different approaches on minimum sample size determination reported in the field of bio-informatics and other clinical studies, to name a few, micro array data [1], cDNA arrays [2] transcription level [3] etc. Based on these works, data-driven hypotheses could be developed which in turn furthers vibration signal analysis research. Though it has been reported in [4,5] for fixing sample size to train the classifier for some particular applications of vibration analysis, the same sample size cannot be used for the present study. Hence, this paper focuses on determination of sample size to build a robust classifier for fault diagnosis. There are many ways available for determination of sample size viz. for tests of continuous variables [6], for tests of proportions [7], for time-to-event (survival) data [8], for receiver operating curve (ROC) analysis [9], for logistic and Poisson regression [10], repeated measurements [11], precision [12], paired samples [13], measurement of agreement [14], and power [15]. Studies were also carried out to discuss issues surrounding estimating variance, sample size re-estimation based on interim data [16], studies with planned interim analyses [17], and ethical issues [18]. However, there are certain issues to be addressed in implementation of such techniques to have better statistical stability.

In machine learning approach, the vibration signals are typically subjected to analyses such as hypothesis testing, classification [19], regression and clustering that rely on statistical parameters to draw conclusions [20–23]. However, these parameters could not be reliably estimated with only a small number of vibration signals. Since the statistical stability of conclusions largely depends on the accuracy of parameters used, a certain minimum number of vibration signals are required to ensure confidence in the sample distribution and accuracy of parameter values. The objective of this paper is to determine the minimum number of samples required to separate the classes with statistical stability using *F*-test based statistical power analysis.

1.1. Methodology

The methodology has been illustrated with the help of a typical automobile hydraulic brake system. Fig. 1 shows the methodology of the proposed study. Referring Fig. 1, the vibration signal under different fault conditions were acquired from the brake setup. Number of descriptive statistical features were extracted from the vibration signal. Among them most important features were selected using decision tree. The minimum number of samples required for classification with statistical stability using *F*-test based statistical power analysis. The minimum sample size is also determined using an entropy based algorithm called ‘C4.5 decision tree’. The results of power analysis are compared with that of C4.5 decision tree algorithm and sample size guidelines are presented for the considered system at the conclusion section.

2. Experimental studies

Experimental study was carried out on a hydraulic brake system setup shown in Fig. 2 [24]. A commercial passenger car's (Maruti Swift) hydraulic brake system (shown in Fig. 2) was used to fabricate the brake test setup. The test rig consists of drum and disc brake coupled together by a shaft. The shaft is in turn run by a DC motor (1HP) coupled to a belt drive system. DC motor consists of an inbuilt drive. A lever is placed at the top of the motor which is connected to the accelerator pedal providing variable speeds up to 2500 rpm. If the accelerator pedal is pressed, the spring attached to the lever is compressed. Hence the pulley shaft is connected with the inbuilt drive. The power is transmitted from the drive to the pulley. When the pedal is released, the pulley shaft gets detached from the drive shaft through spring expansion. Brake pedal is provided to the left side of the accelerator pedal. It is attached to the piston in the master cylinder via a push rod. Master cylinder, the most important part of hydraulic brake is provided with pistons to move along the bore. Since hydraulic brakes are prominent brake system in medium motor vehicle like cars, in order to experiment with the components used in real world, branded vehicles (cars)

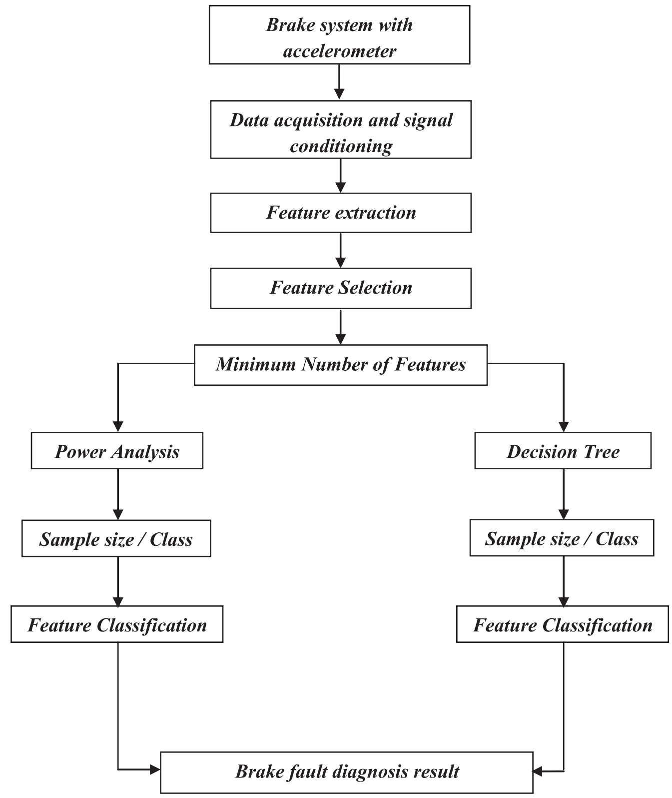


Fig. 1. Flow chart – methodology.

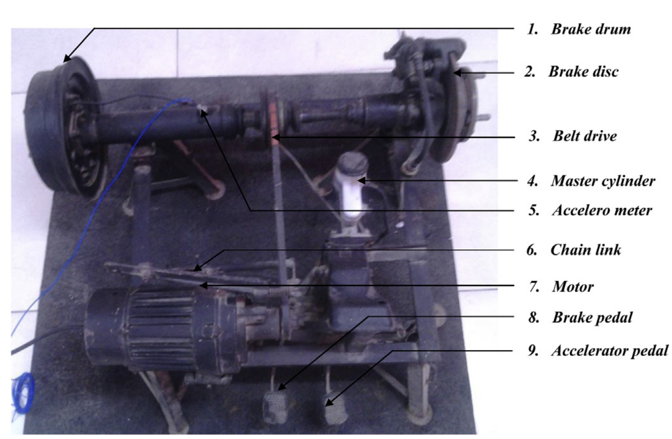


Fig. 2. Brake fault diagnosis – experimental setup.

parts were considered. The dimension of test rig is 80 cm × 80 cm × 40 cm.

Piezoelectric type accelerometer was used as transducer for acquiring vibration signals. An uni-axial accelerometer of 50 g range, 100 mV/g sensitivity, and resonant frequency around 40 Hz was used. The DAQ system is NI USB 4432 model. The vibration signals are acquired using data acquisition system with the sampling frequency of 24 kHz. The sample length was assumed to be 1024 by using Nyquist sampling theorem. 55 samples for each condition were taken from the hydraulic braking system.

Initially the test rig was assumed to be in good condition. (All components were brand new). The frequently occurred nine most important fault conditions were simulated for testing. They are, air in the brake fluid (AIR), brake oil spill on disc brake (BOS), drum brake pad wear (DRPW), disc brake pad wear (even)-inner (DPWI), disc brake pad wear (even)-inner and outer (DPWIO), disc brake pad wear (uneven-inner) (UDPWI), disc brake pad wear (uneven)-inner and outer (UDPWIO), reservoir leak (RL), drum brake mechanical fade (DRMF). The vibration signals were measured from the hydraulic brake system working under constant braking condition (original speed 667 rpm, brake load 67.7 N). From the accelerometer, the vibration signals for different fault conditions were taken with the following settings.

A LabVIEW graphical program was used to store the signal in the computer. The digital version of the signal was then processed to extract different feature which contains information's that are

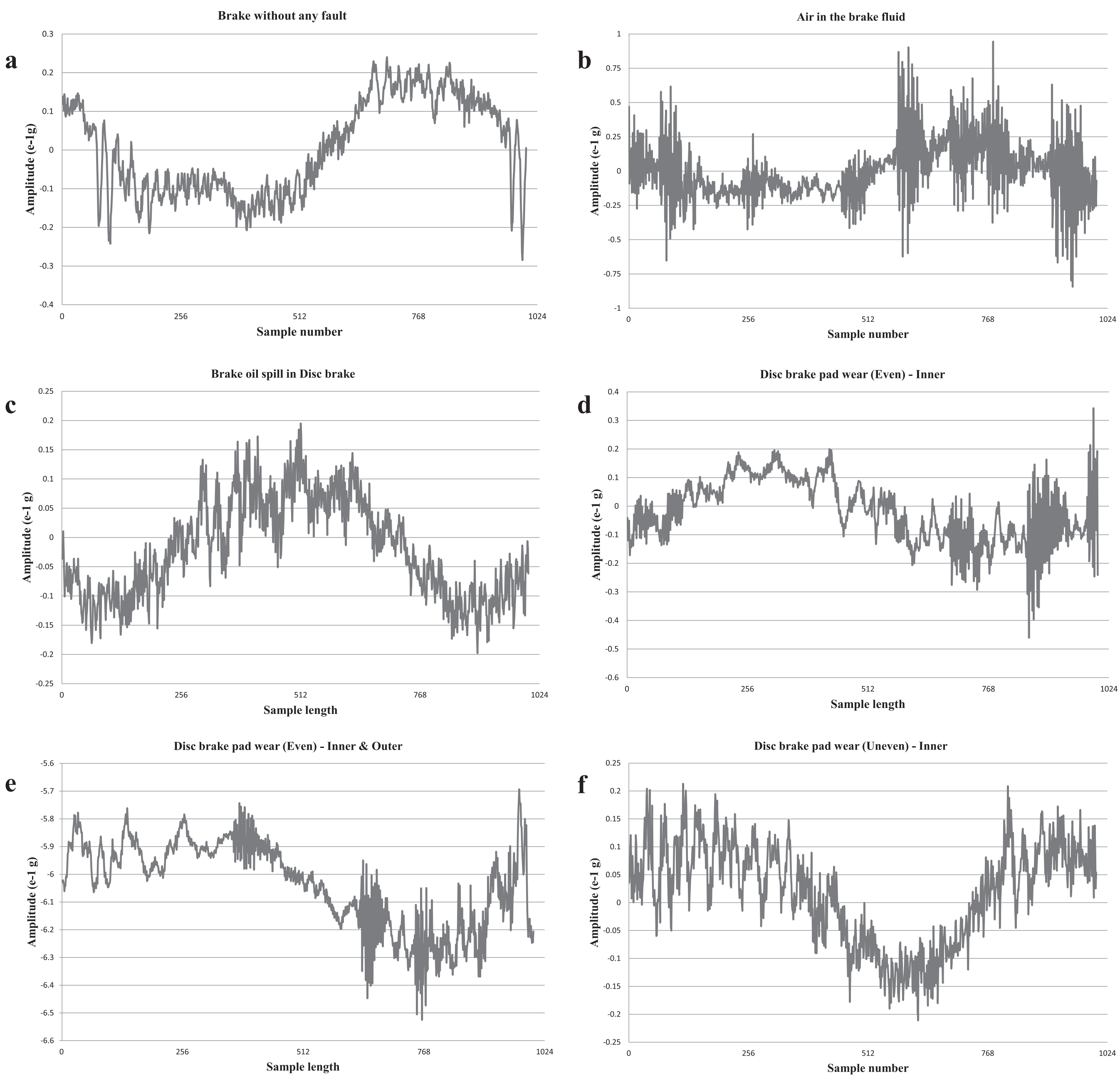


Fig. 3. (a) Vibration signal – brake without any fault. (b) Vibration signal – air in the brake fluid. (c) Vibration signal – brake oil spill. (d) Vibration signal – disc brake pad wear (even) – inner. (e) Vibration signal – disc brake pad wear (even) – inner & outer. (f) Vibration signal – disc brake pad wear (uneven) – inner. (g) Vibration signal – disc brake pad wear (uneven) – inner & outer. (h) Vibration signal – drum brake mechanical fade. (i) Vibration signal – drum brake pad wear. (j) Vibration signal – reservoir leak.

relevant to the considered fault conditions. Fig. 3(a)–(j) shows the time domain signals taken from the brake setup. Once the faults were simulated, the vibration signals were recorded and feature extraction and feature selection was carried out using these vibration signals [24].

3. Feature extraction and feature selection

The process of computing some specific measures that represent the vibration signals is called feature extraction. A fairly wide set of statistical parameters are selected as the basis of the study. They are mean, standard error, sample variance, kurtosis, skewness, minimum, maximum, standard deviation, count, mode and median. These parameters are called statistical features. The statistical parameters can be calculated using the formulas provided in Table 1. The statistical information contained in the signals has been extracted using a visual Basic soft tool with Microsoft Office Excel. The statistical features given in Table 1 have been extracted from the raw vibration signal using a statistical tool in Microsoft Office Excel. The process of extracting statistical features was described for bearing fault diagnosis by Sugumaran et al. [23]. Following the footsteps of Sugumaran et al. feature extraction was carried out.

There are many techniques available for feature selection. The commonly used techniques for selection of features are principal component analysis (PCA), genetic algorithm (GA), and decision tree (DT) [23]. In a study by Sugumaran et al. the use of a decision tree to identify the best feature selection from a given set of samples for classification was illustrated [23]. The most important feature will be placed on top of the decision tree and others will follow. Based on this, the most important features were identified namely minimum, standard error, sample variance, kurtosis and skewness.

Table 1
Statistical features definition.

Name of the statistical features	Formula/description
Standard error	$\sqrt{\frac{1}{n-2}\left[\sum(y-\bar{y})^2 - \frac{\sum[(x-\bar{x})(y-\bar{y})]^2}{\sum(x-\bar{x})^2}\right]}$
Standard deviation	$\sqrt{\frac{n\sum x^2-(\sum x)^2}{n(n-1)}}$
Sample variance	$\sqrt{\frac{n\sum x^2-(\sum x)^2}{n(n-1)}}$
Kurtosis	$\left\{\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i-\bar{x}}{S_d}\right)^4\right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$
Skewness	$\frac{n}{(n-1)(n-2)}\sum\left(\frac{x_i-\bar{x}}{S_d}\right)^3$
Maximum value	Maximum signal point value in a given signal.
Minimum value	Minimum signal point value in a given signal.
Range	Difference in maximum and minimum signal point values for a given signal.
Sum	Sum of all feature values for each sample.
Mean	The arithmetic average of a set of values or distribution.
Median	Middle value separating the greater and lesser halves of a data set.
Mode	A statistical term that refers to the most frequently occurring number found in a set of numbers. (i.e.) The mode is the value that appears most often in a set of data. In other words, it is the value that is most likely to be sampled.

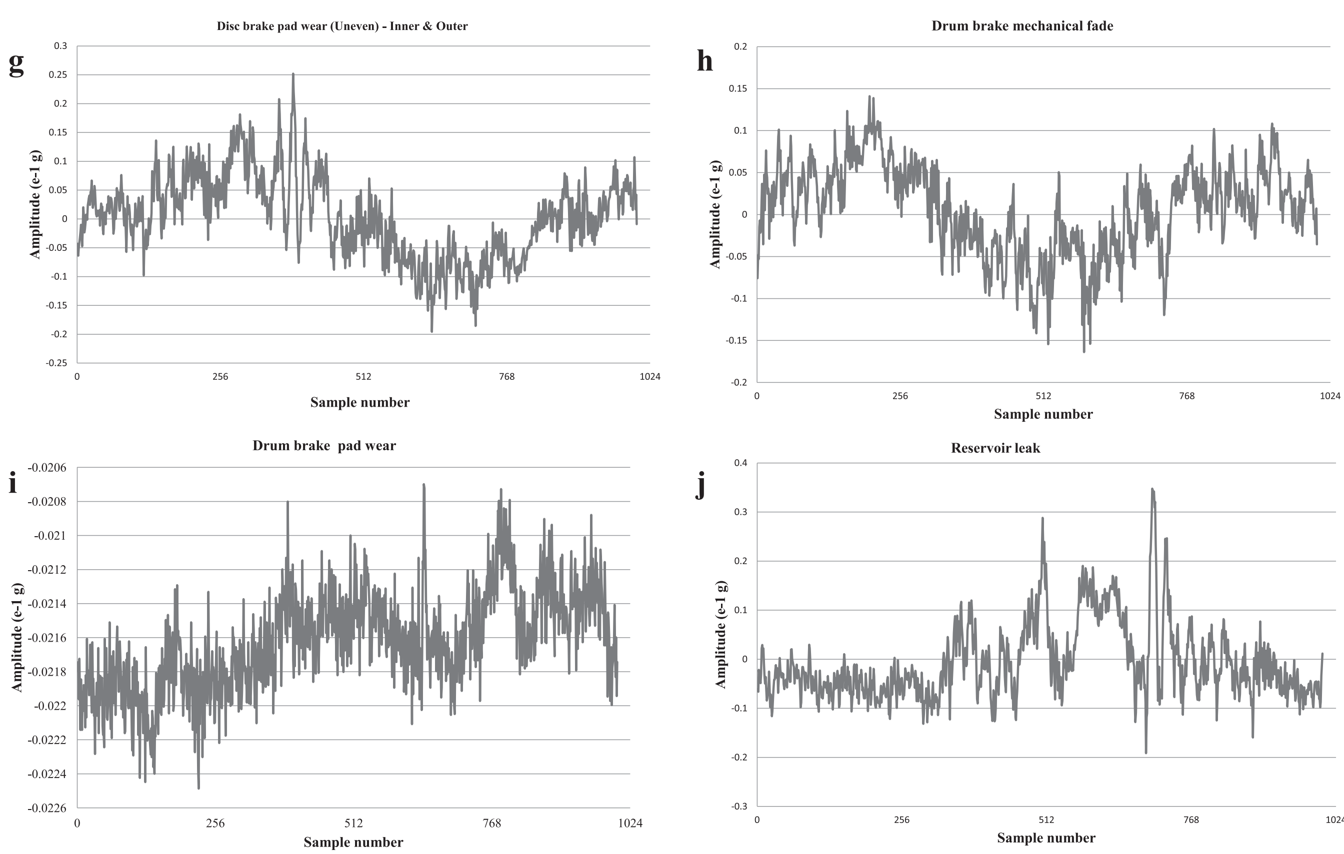


Fig. 3. (*continued*).

4. C4.5 decision tree algorithm

Fault diagnosis can be viewed as a data mining problem where one extracts information from the acquired data through a classification process. A predictive model for classification invokes the idea of branches and trees identified through a logical process. Any good classifier should have the following properties:

1) It should have good predictive accuracy; it is the ability of the model to correctly predict the class label of new or previously unseen data.
2) It should have good speed.
3) The computational cost involved in generating and using the model should be as low as possible.
4) It should be robust; Robustness is the ability of the model to make correct predictions given the noisy data or data with missing values.
5) The level of understanding and insight that is provided by classification model should be high enough.

It is reported that C4.5 model introduced by J.R. Quinlan satisfies with the above criteria and hence the same is used in the present study [25]. The classification is done through a decision tree with its leaves representing the different fault conditions of the brake system. Fig. 2 shows the decision tree obtained from C4.5 decision tree algorithm. The sequential branching process ending up with the leaves here is based on conditional probabilities associated with individual features. Decision tree algorithm (C4.5) has two phases: building and pruning. The building phase is also called as 'growing phase' [26]. As is customary the samples are divided into two parts: training set and testing set. Training set is used to train classifier and testing set is used to test the validity of the classifier. 10-fold cross-validation has been employed to evaluate classification accuracy.

The training process of C4.5 using the samples with continuous-valued attributes is as follows.

1) The tree starts as a single node representing the training samples.
2) If the samples are all of the same class, then the node becomes a leaf and is labeled with the class.
3) Otherwise, the algorithm discretises every attribute to select the optimal threshold and uses the entropy-based measure called information gain (discussed in Section 5.1) as heuristic for selecting the attribute that will best separate the samples into individual classes.
4) A branch is created for each best discrete interval of the test attribute, and the samples are partitioned accordingly.
5) The algorithm uses the same process recursively to form a decision tree for the samples at each partition.
6) The recursive partitioning stops only when one of the following conditions is true:
   (a) All the samples for a given node belong to the same class or
   (b) There are no remaining attributes on which the samples may be further partitioned.
   (c) There are no samples for the branch test attribute. In this case, a leaf is created with the majority class in samples.
7) A pessimistic error pruning method is used to prune the grown tree to improve its robustness and accuracy.

5. Determination of sample size

As discussed in earlier sections, determining minimum sample size of vibration signals to train the classifier are very essential for a categorization problem of different conditions for a machine using vibration signals. In the present study, the eleven measures obtained from vibration signals are served as features. These features were used as representatives of the vibration signals instead of using them directly. This has been done only to reduce the dimension of the problem to a sizable number. Further dimensionality reduction was carried out using C4.5 algorithm to eliminate less contributing features from the feature set. The method and procedure of performing the same using C4.5 algorithm is explained in Section 5. In power analysis, the selected five features along with their class label have been used as a data set. The method of performing power analysis is discussed in Section 5.1. The results obtained in this method were verified with the help of a functional test using C4.5 algorithm. As C4.5 decision tree algorithm can be used as a classifier with 10-fold cross-validation method, the number of samples is decreased from 55 per class to five per class. The results are presented and discussed in Section 6.

5.1. Power analysis

Sample size has a great influence in any experimental study, because the result of an experiment is based on the sample size. Power analysis has been used in many applications [4,5]. Performing power analysis and sample size estimation is an important aspect of experimental design, because without these calculations, sample size may be too high or too low. If sample size is too low, the experiment will lack the precision to provide reliable answers to the questions that are investigated. If sample size is too large, time and resources will be wasted, often for minimal gain. Power analysis has been used in many applications.

In power analysis, the test family has been chosen as F-test under the statistical test MANOVA with repeated measures within between interactions. It is based upon two measures of statistical reliability, namely the confidence interval $(1-\alpha)$ and power $(1-\beta)$. As it is a hypothesis test, the test compares null hypothesis (H_0) against the alternative hypothesis (H_1). The null hypothesis is assumed to be the means of the classes are the same whereas alternative hypothesis H_1 is defined to be the means of classes are not same while the confidence level of a test is the probability of accepting null hypothesis, the power of a test is the probability of accepting the alternative hypothesis [1]. Alternatively false positives, α (Type I error) is the probability of accepting alternative hypothesis while false negatives, β (Type II error) is the probability of accepting the null hypothesis.

The estimation of sample size in power analysis is done such that the confidence and the power (statistical reliability measures) in hypothesis test can reach predefined values. Typical analyses may require the confidence of 95% and the power of 95%. The confidence level and the power are calculated from the distributions of the null hypothesis and alternative hypothesis. Defining these distributions depends on the statistical measures being used in the hypothesis test. For a two class problem, the method and procedure of computing hypothesis test using t-distribution is explained in [1]. In case of multi-class problem (number of classes greater than two), instead of T-statistic, the F-statistic measure derived from Pillai's V formula [26] is used for the estimation of sample size. Pillai's V is the trace of the matrix defined by the ratio of between-group variance (B) to total variance (T). It is a statistical measure often used in multivariate analysis of variance (MANOVA) [26]. The Pillai's V trace is given by

$$V = \text{trace}\left(BT^{-1}\right) = \sum_{i=1}^{h} \frac{\lambda_i}{\lambda_i + 1} \quad (1)$$

where λ_i is the ith eigen value of $W^{-1}B$ in which W is the within-group variance and h is the number of factors being considered in MANOVA, defined by $h = c - 1$. A high Pillai's V means a high amount of separation between the samples of classes, with the between-group variance being relatively large compared to the total variance. The hypothesis test can be designed as follows using F statistic transformed from Pillai's V.

$$H_0: \mu_1 = \mu_2 = \mu_3 \ldots = \mu_c; \; H_1 : \text{Exists under the condition } \mu_i - \mu_j \neq 0 \tag{2}$$

$$H_0: F = \frac{(V/s)/(ph)}{(1-(V/s))/[s(N-c-p+s)]} \sim F(ph, s(N-c-p+s)) \tag{3}$$

$$H_1: F = \frac{(V/s)/(ph)}{(1-(V/s))/[s(N-c-p+s)]} \sim F[ph, s(N-c-p+s), \Delta = s\Delta_e N] \tag{4}$$

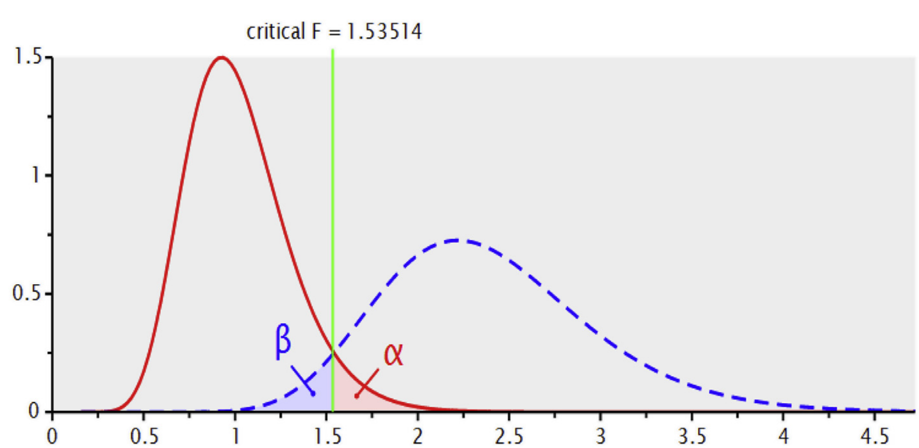
with $\Delta_e = \dfrac{V_{crit}}{(s - V_{crit})}$,

Where p and c are the number of variables and the number of classes, respectively. s is defined by min(p, h). By using these defined distributions of H_0 and H_1, the confidence level and the power can be calculated for a given sample size and effect size. The method used for multi-class problem is used here to estimate the minimum sample size for statistical stability whereby the sample size is increased until the calculated power reaches the predefined threshold of $(1 - \beta)$. However, here is a limitation that the value of p cannot be larger than $N - c + s = N - 1$. This analysis may produce a misleading sample size estimate when the real data set is not consistent with the assumption (normality and equal variance) underlying the statistic used in power analysis. To check the effect of possible violations of the assumptions on the estimated sample size, the actual power and mean differences between classes are compared to the predefined values. The actual values in both cases studied were sufficiently large that we need not be worried about the impact of data which does not perfectly match the normality or equal variance assumptions.

6. Results and discussion

In this study, vibration signals of a good and nine different faulty conditions of a hydraulic brake system have been considered. The statistical power analysis F-test was employed to find the minimum number of samples required to train the classifier with statistical stability so as to get good classification accuracy. A data set consisting of 550 samples has been considered. The null hypothesis, H_0 was assumed to be the means of the classes are the same whereas the alternate hypothesis H_1 was assumed to be the means of the classes are not the same. In statistics, type I error α is rejecting null hypothesis when it is true and type II error β is accepting null hypothesis when it is false. The power of the statistical test is the probability that the test will reject null hypothesis when it is false. That is, probability of not committing type II error and therefore power is denoted by $(1 - \beta)$. At first, the power level $(1 - \beta)$ and the confidence interval $(1 - \alpha)$ were assumed to be 99%. In this power analysis test, the sample size has been obtained for the given power level, α error probability, repetitions and number of groups. Further, the sample size has been obtained for various power levels 95%, 90%, 85%, 80%, and 75% with α error probability 5%, 10%, 15%, 20% and 25% respectively. The central and non-central distribution of the data set in power analysis is shown in Fig. 4.

From the mean and covariance matrix of the given data set, the effect size and the corresponding value of the Pillai's V parameter were found to be 0.5876 and 1.0267 respectively. One can easily measure the statistical stability of the data set using the value of Pillai's V parameter. If Pillai's V value is greater than 0.5, then the statistical stability will be more and the required number of samples to be trained will be minimum and vice-versa. Sample size has been obtained as 51 for all ten classes for the predefined power level (99%) and α error probability (1%) with this Pillai's V value. Hence it reduces to 5 per class. That is, it is enough if one has 5 samples to train the classifier to get good classification accuracy and to retain a power level of 99%. Also, one can find the required sample size with the power levels nearer to 99% from Fig. 5. At this point, it is more appropriate to ask about the number of samples required if the test can accommodate less than 99% power level namely, 95%, 90%, 85%, 80% and 75%. Fig. 6 shows the number of samples required for the given α error probability and for various power level from 75% to 95% in steps of 5%. It can also be plotted as shown in Fig. 7. That is, the sample size as a function of power level for various α error probability from 5% to 25% in steps of 5%. The



Sample size calculation with effect size f(V) = 0.5876299, a error probability = 0.01,
power (1 _ β error probability) = 0.95, Number of groups = 10 and repetitions = 5 with Pilli's V formula.

Fig. 4. *F*-Tests – MANOVA: repeated measures.

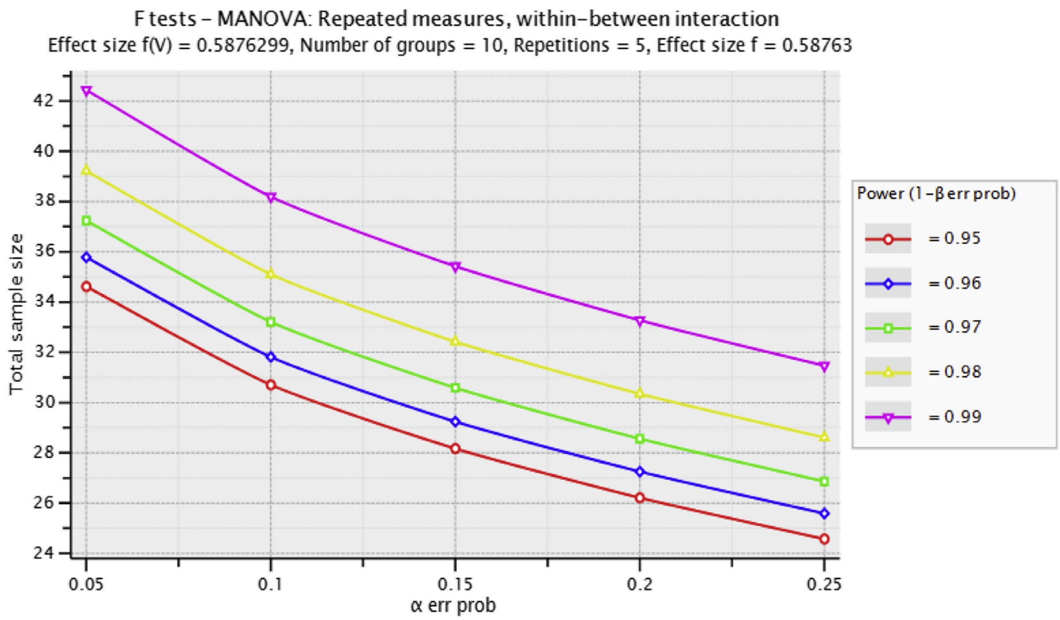


Fig. 5. Total sample size as a function of α error probability for various power level.

sample size for different pairs of power level $(1-\beta)$ and α error probability is also presented in Table 2.

Sample size is not only the function of power level and α error probability in power analysis but also a function of effect size. Effect size measures play an important role in statistics. In statistics, an effect size is a measure of the strength of the relationship between two variables in a statistical population or a sample-based estimate of that quantity. A set of experiments were carried out to find the sample size as a function of effect size. Sample size as a function of effect size for various α error probabilities and for various power levels $(1-\beta)$ have been shown in Figs. 8 and 9 respectively. From these figures, it is evident that the sample size decreases if either the effect size increases or the α error probability increases. As effect size increases the total sample size decreases for a given power level. As power level increases, the required sample size also increases with increases in effect size. Finally, to study the influence of effect size on α error probability, a set of experiments was carried out for different power levels and the results are shown in Fig. 10. As α error probability increases the effect size decreases. For a given power level, the effect size decreases as α error probability increases.

The results obtained using power analysis has been validated through a functional test namely, C4.5 decision tree algorithm. Due to easiness in training, general classification accuracy and computation complexity, C4.5 algorithm has been chosen amongst the pool of classifiers. In this study, 55 samples from each class have been considered and all the samples have been used for training the classifier with 10-fold cross validation method. Here, the classifier's job is a two phase process namely, feature selection and feature classification. A decision tree (Fig. 11) has been generated using this

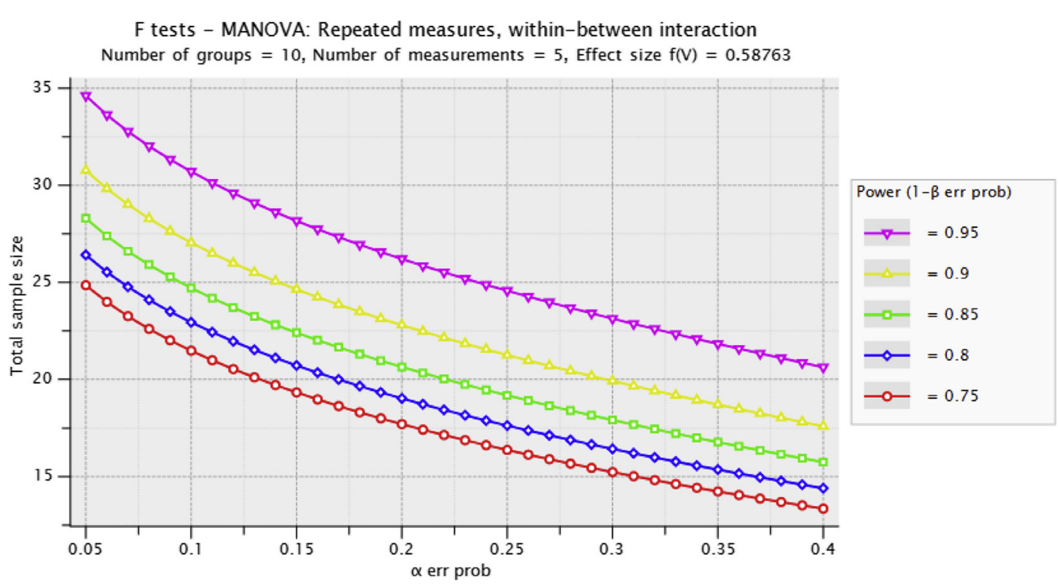


Fig. 6. Total sample size as a function of α error probability for various power level.

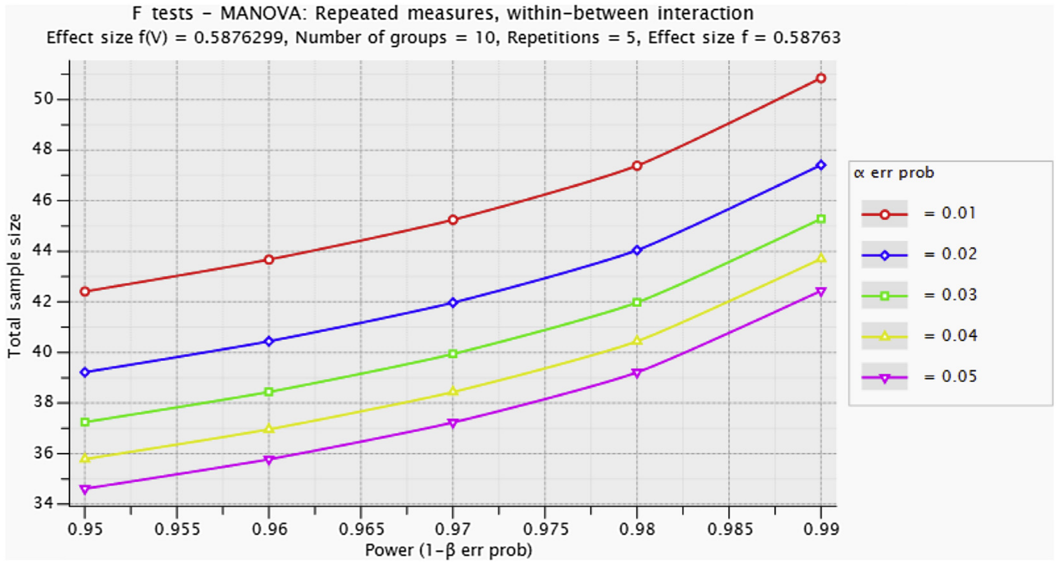


Fig. 7. Total sample size as a function of power level for various α error probability.

Table 2
Power analysis test results.

Output parameter	$\alpha = 0.01$, $1-\beta = 0.99$	$\alpha = 0.05$, $1-\beta = 0.95$	$\alpha = 0.10$, $1-\beta = 0.90$	$\alpha = 0.15$, $1-\beta = 0.85$	$\alpha = 0.20$, $1-\beta = 0.80$	$\alpha = 0.25$, $1-\beta = 0.75$
Non centrality parameter λ	70.443015	48.343246	38.6746	31.76471	27.62471	26.24347
Critical *F*	1.7509	1.5352	1.4278	1.3655	1.3141	1.2544
Denominator d*f*	164	100	72	52	40	36
Total sample size for 4 classes	51	35	28	23	20	19
Sample size per classes	≈5	≈4	≈3	≈2	≈2	≈2
Actual power	0.990305	0.9535217	0.915947	0.864861	0.831498	0.8447445

Power analysis test results with effect size $f(V) = 0.5876299$, number of groups = 10, repetitions = 5 with Pilli's *V* formula value = 1.0267052, numerator d*f* = 36.0000.

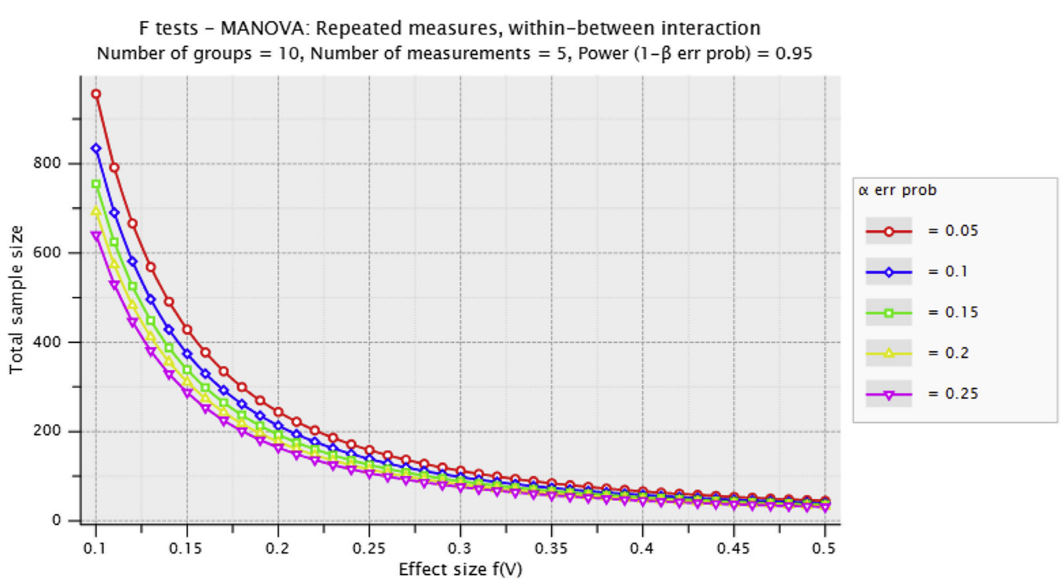


Fig. 8. Total sample size as a function of effect size for various α error probability.

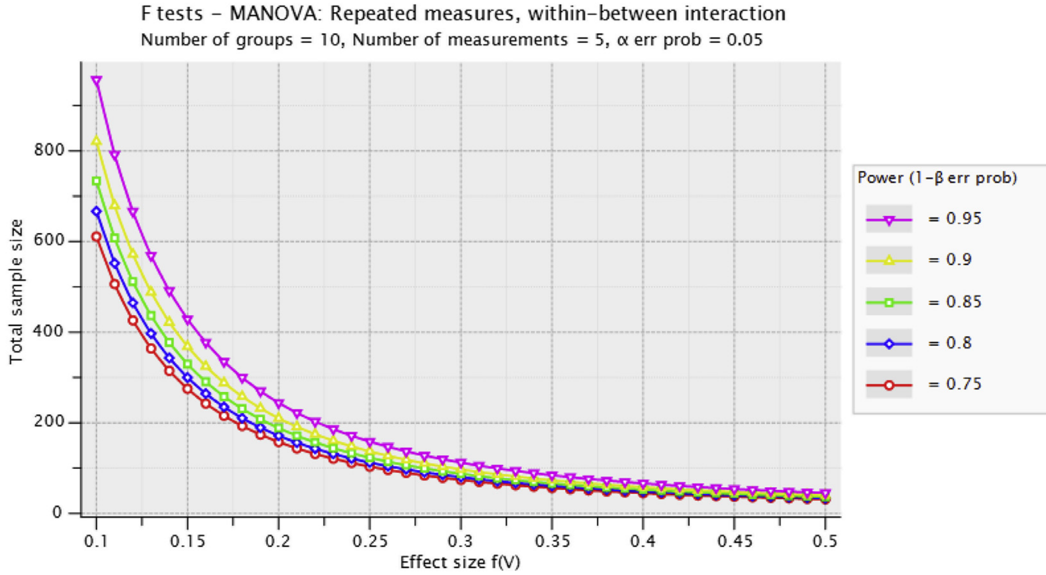


Fig. 9. Total sample size as a function of effect size for various power level.

classifier and the five top most features namely have been chosen as the selected features for the further study. To examine the variation in the classification accuracy due to the sample size, five samples from each class have been reduced and the classification accuracy has been noted. This process has been carried out till it remains to have five samples in each class. These results have been plotted and shown in Fig. 12. From Fig. 12, one can observe that the classification accuracy goes down significantly when the sample size is reduced to below five. Also, the classification accuracy increases when the sample size increases. It shows that it is enough if one has five samples per class to train the classifier so as to get good classification accuracy. However, present study aims to find the sample size for training the classifier with statistical stability. Hence, mean absolute error and the root mean squared error as a function of sample size have been considered and shown in Figs. 13 and 14 respectively. For comparing the results obtained in power analysis and C4.5 decision tree algorithm, a representative value corresponding to 5% of α error probability was taken from Table 2. As per power analysis result, from Table 2, if one can accommodate 5% of α error probability and willing to accept 95% of power level, then for given data set the minimum required sample size is 5. This means, if 5 samples were used for training the classifier, the maximum α error probability that likely to happen would be 5%. This has to be validated with C4.5 algorithm results. From Fig. 13, it is evident that the mean absolute error is below 10% (i.e. it did not exceed 10%) for cases whose sample size is greater than or equal to 10. The corresponding root mean squared error is shown in Fig. 14.

In classification problem, the mean absolute error of the classifier is a measure of type I error (α error probability). Type I error is an error due to misclassification of the classifier. The mean absolute error is a measure of type one error in classification problem. In general, type I error is rejecting a null hypothesis when it is true. α

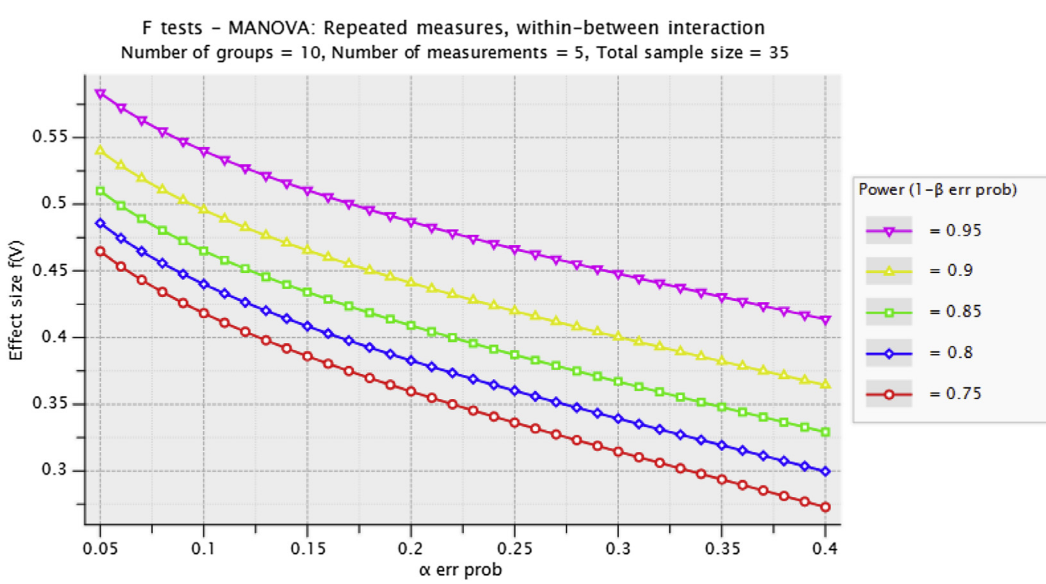


Fig. 10. Effect size as a function of α error probability for various power level.

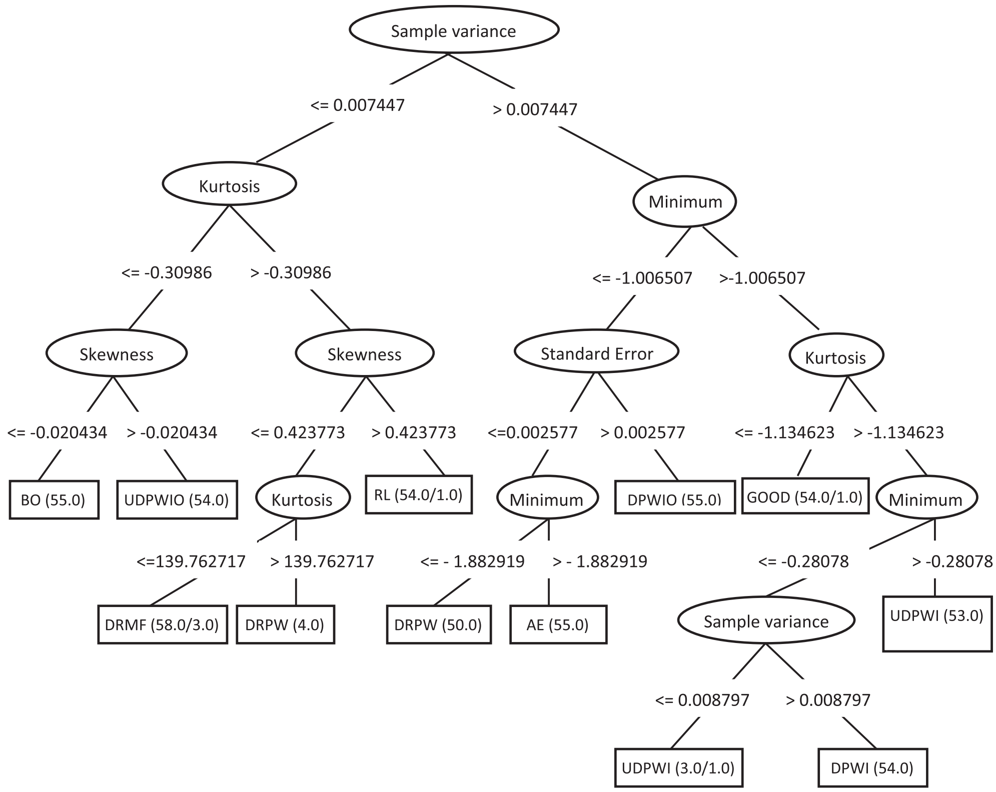


Fig. 11. Decision tree.

error probability is a measure of type I error in hypothesis testing and hence, the equivalence is obvious. From the above discussion, the results of power analysis are true and the actual error did not exceed the upper bound (5%) found in power analysis. A similar exercise of validating the results at other points also assures the validity of the power analysis test. Thus one can confidently use the sample size suggested by power analysis for machine learning approach to fault diagnosis of automobile hydraulic brake system.

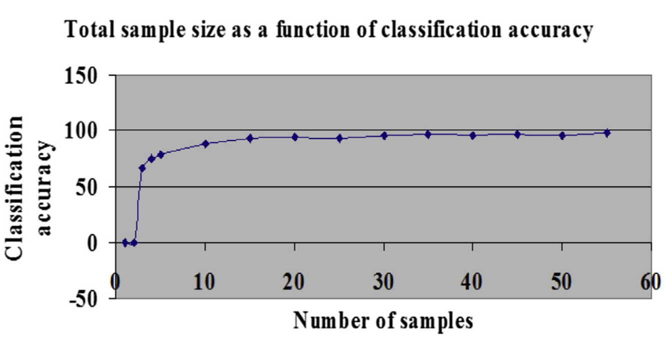


Fig. 12. Sample size as a function of classification accuracy.

7. Conclusion

In this study, an automobile hydraulic brake system has been considered with a good and nine different faulty conditions. A statistical method called power analysis has been used to find the minimum sample size to train the classifier so as to get good classification accuracy with statistical stability. The statistical features have been used for finding the sample size. The sample size for 99% power level has been obtained as five per class. Sample size for various α error probability and power level are also presented in Table 1. The results obtained using power analysis has been validated by using a functional test namely, C4.5 decision tree algorithm. The results and the graphs presented in Section 6 will serve as a guideline for fixing the sample size in fault diagnosis of an automobile hydraulic brake system.

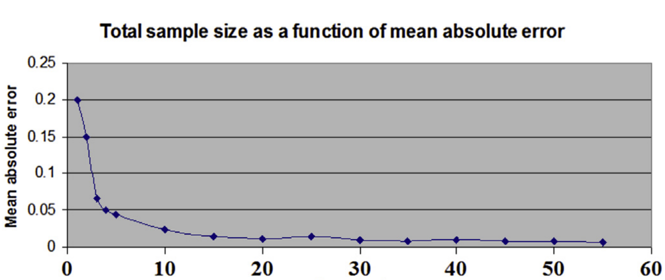


Fig. 13. Sample size as a function of mean absolute error.

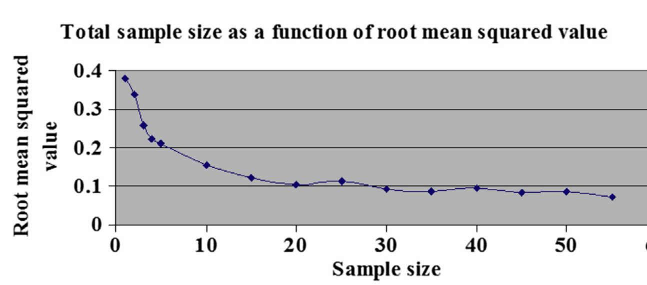


Fig. 14. Sample size as a function of root mean squared value.

References

[1] Daehee Hwang, William A. Schmitt, George Stephanopoulos, Gregory Stephanopoulos, Determination of sample size and discriminatory expression patterns in microarray data, Bioinformatics 18 (2002) 1184–1193.
[2] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray, Science 270 (1995) 467–470.
[3] D.J. Lockhart, et al., Expression monitoring by hybridization to high density oligonucleotide arrays, Nat. Biotechnol. 14 (1996) 1675–1680.
[4] V. Indira, R. Vasantha Kumari, V. Sugumaran, Minimum Sample Size Determination of Vibration Signals in Machine Learning Approach to Fault Diagnosis Using Power Analysis, Expert Systems with Applications, vol. 37, Elsevier Journal Publications, 2010, pp. 8650–8658.
[5] V. Indira, R. Vasantha Kumari, N.R. Sakthivel, V. Sugumaran, A Method for Calculation of Optimum Data Size and Bin Size of Histogram Features in Fault Diagnosis of Mono Block Centrifugal Pump, Expert Systems with Applications, vol. 38, Elsevier Journal Publications, 2011, pp. 7708–7717.
[6] S.J. Day, D.F. Graham, Sample size estimation for comparing two or more treatment groups in clinical trials, Stat. Med. 10 (1991) 33–43.
[7] P.A. Feigl, Graphical aid for determining sample size when comparing two independent proportions, Biometrics 34 (1978) 111–122.
[8] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29–36.
[9] N.A. Obuchowski, Computing sample size for receiver operating characteristic studies, Am. J. Roentgenol. 175 (2000) 603–608.
[10] S.B. Bull, Sample size and power determination for a binary outcome and an ordinal exposure when logistic regression analysis is planned, Am. J. Epidemiol. 137 (1993) 676–684.
[11] S. Greenland, On sample-size and power calculations for studies using confidence intervals, Am. J. Epidemiol. 128 (1988) 231–237.
[12] S.L. Beal, Sample size determination for confidence intervals on the population mean and on the difference between two population means, Biometrics 45 (1989) 969–977.
[13] A. Donner, M. Eliasziw, A goodness-of-fit approach to inference procedures for the kappa statistic: confidence interval construction, significance-testing and sample size estimation, Stat. Med. 11 (1992) 1511–1519.
[14] M.A. Birkett, S.J. Day, Internal pilot studies for estimating sample size, Stat. Med. 13 (1994) 2455–2463.
[15] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, Axel Buchner, G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Methods 39 (2) (2007) 175–191.
[16] R.H. Browne, On the use of a pilot sample for sample size determination, Stat. Med. 14 (1995) 1933–1940.
[17] N.L. Geller, S.J. Pocock, Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners, Biometrics 43 (1987) 213–223.
[18] J.D. Lantos, Sample size: profound implications of mundane calculations, Pediatrics 91 (1993) 155–157.
[19] V. Sugumaran, G.R. Sabareesh, K.I. Ramachandran, Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine, Expert Syst. Appl. 34 (2008) 3090–3098.
[20] S.N. Kavuri, V. Venkatasubramanian, Using fuzzy clustering with ellipsoidal units in neural networks for robust fault classification, Comput. Chem. Eng. 17 (8) (1993) 765–784.
[21] R. Vaidyanathan, V. Venkatasubramanian, Representing and diagnosing dynamic process data using neural networks, Eng. Appl. Artif. Intell. 5 (1) (1992) 11–21.
[22] W.J. Wang, P.D. McFadden, Early detection of gear failure by vibration analysis—I. Calculation of the time–frequency distribution, Mech. Syst. Signal Process. 7 (3) (1993) 193–203.
[23] V. Sugumaran, K.I. Ramachandran, Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing, Mech. Syst. Signal Process. 211 (2007) 2237–2247.
[24] R. Jegadeeshwaran, V. Sugumaran, Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical feature, Measurement 46 (9) (2013) 3247–3260.
[25] J.R. Quinlan, Decision trees and multi-valued attributes, Machine Intell. 11 (1985) 81–106.
[26] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, second ed., Erlbaum, Hillsdale, NJ, 1988.