

Induction motor bearing fault diagnostics using i-kazTM and decision tree classification

M.S. Othman^{1*}, M.Z. Nuawi¹ and R. Mohamed²

¹ Department of Mechanical and Material Engineering,

² Department of Electrical, Electronic and Systems Engineering,
Universiti Kebangsaan Malaysia, Jalan Bangi
43600 UKM Bangi, Selangor, Malaysia.

*Email: sufian@jkr.gov.my

Phone: +6016-997 1021; Fax: +603-8911 8389

ABSTRACT

In this paper, the I-kazTM method is proposed for the detection of induction motor bearing faults using vibration signals, which has not been presented so far. The purpose of the study is to compare this new technique with the classical kurtosis method in the time domain and to validate the performance of the proposed I-kaz coefficient using a decision tree classification. Three bearing conditions are investigated; i.e. normal, ball fault and inner race fault; with a very small fault size (0.1778 mm). All faulty bearings are artificially damaged using electro-discharge machining and placed in the motor drive end side. The experimental test rig consists of a 2 HP induction motor, a torque transducer, a dynamometer and control electronics. Vibration data is obtained using an accelerometer and analyzed using MATLAB software for the time domain analysis which include the I-kaz graphical and coefficient comparison with time waveform and kurtosis values for all bearing conditions. Then, both features are used to train a conditional inference tree (CTree) fault classifier separately. The proposed I-kaz coefficient provides higher percentage differences between all faulty and normal bearings compared to kurtosis. However, the I-kaz graphic presents similar identification as the time waveform where only the inner race fault is distinguished from the normal bearing. The training classification results also revealed that the I-kaz coefficient is significantly better with an accuracy of 99.64% and a Kappa value of 0.9946 compared to kurtosis at only 63.57% and 0.4536, respectively. Furthermore, all test data were classified accordingly using the I-kaz coefficient whereas for kurtosis, only 65% are correctly classified with the 0.475 Kappa value. It is proved that the I-kazTM method is suitable for induction bearing fault detection and is recommended as a classification feature, especially for the diagnostics of ball fault which is the most difficult to diagnose.

Keywords: Condition monitoring, decision tree classification, bearing fault diagnosis, vibration signal, I-kazTM.

INTRODUCTION

Bearings are among the most common components in rotating machinery such as electrical motors, gearboxes, pumps, fans, etc. As a consequence, bearing fault is also one of the primary causes of failure in rotating equipment [1]. Therefore, bearing fault detection is important in order to prevent an abnormal event which can lead to

productivity loss, emergency breakdown, and catastrophic damage. The condition monitoring of bearing can be done using many techniques. Toliyat et. al. [2] specified that the sources of signal-based electric motor fault diagnosis can be from vibrations, shock pulses, temperature, acoustic noises, electromagnetic fields, output power variations, gas, oil, radio-frequency, motor current, etc. Vibration signals is a popular tool in fault diagnosis and has been applied successfully by many researchers [3-7].

Bearing failure can be categorized as ‘distributed’ or ‘localized’ [8]. Distributed defects can happen during the manufacturing process, improper installation or abrasive wear. This type of defect includes surface roughness, waviness, misaligned races, and off size balls. Whereas, localized defects include cracks, pits, and spalls on the rolling surfaces which might be caused by fatigue failure due to overloading or shock loading to the bearings during operation and installation. Jin et al. [9] listed common causes of bearing failures which include overloading, contamination, improper lubrication and misalignment.

The time domain analysis is the easiest technique performed directly on the signal time waveform itself. Traditionally, time domain analysis calculates characteristic features as descriptive statistics such as mean, peak, peak-to-peak interval, standard deviation, crest factor, high-order statistics: RMS, skewness, kurtosis, etc. [10]. Kurtosis is one of an earliest methods used for bearing fault detection, first introduced by Dyer & Stewart [11]. Kurtosis, Kur for N number of data (y_1, y_2, \dots, y_n), can be calculated as:

$$Kur = \frac{\frac{1}{N} \sum_{n=1}^N (y_n - \mu)^4}{s^4}, \quad (1)$$

where μ is the mean and s is the standard deviation of data. To have a good bearing condition, kurtosis value must be close to 3 in compliance with the Gaussian distribution, whereas for faulty bearings, kurtosis is relatively higher than 3. However, in some cases when the fault is well advanced, the kurtosis value was reported to go down close to 3. Kurtosis also approaches 3 for small ball faults as disclosed by Smith and Randall [12] at between 3.04 to 3.15, compared to normal bearings which are from 2.76 to 2.96.

The present paper emphasizes on the detection of bearing faults using the new kurtosis-based method called I-kaz™ in exchange for the kurtosis technique used in time-domain analysis. Both features are also used to train the decision tree model separately and are evaluated with the same test data set using accuracy and Kappa coefficient comparison. The study is aimed to compare the performance of I-kaz and kurtosis in bearing fault diagnosis, especially for the detection of ball faults which is certainly the most difficult to diagnose [12].

Integrated Kurtosis-Based Algorithm for Z-Filter (I-kaz™) Method

The I-kaz™ method was developed by Nuawi et al. [13] based on the concept of data distribution about its center points. This method applies both descriptive and inferential statistics. The descriptive part is a numerical value called I-kaz coefficient, Z^∞ and the inferential part is a three-dimensional graphic summarizing data distribution with its low, high, and very high-frequency ranges represented in the x-axis, y-axis, and z-axis respectively. Details of the signal decomposition process are shown in Figure 1, as follows:

- Low-frequency range (LF): 0 to $0.25 f_{max}$,
 - High-frequency range (HF): $0.25 f_{max}$ to $0.5 f_{max}$,
 - Very high-frequency range (VF): $0.5 f_{max}$ to f_{max} ,
- where f_{max} is half of the value of the data sampling rate.

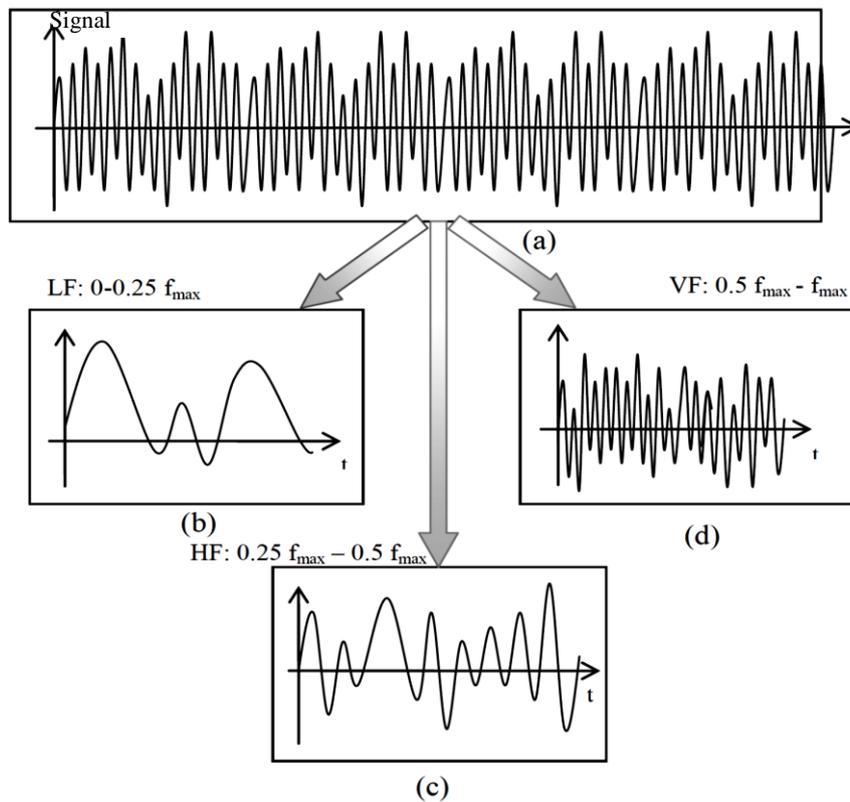


Figure 1. Details of the I-kaz signal decomposition process [13]

Previously, the I-kaz™ method has been successfully applied for the analysis of car engine bearings [14], engine blocks [15], suspension systems [16], machine cutting tools [17], etc. To the author’s best knowledge, this method has never been applied in the field of induction motor bearing diagnostics.

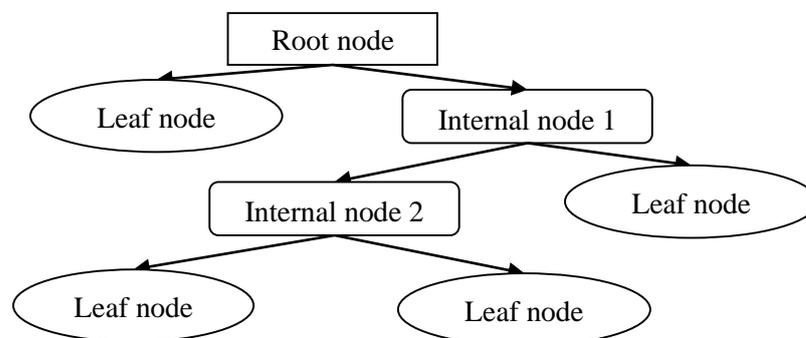


Figure 2. Typical decision tree structure.

Fault Diagnosis Using Decision Tree

Bearing fault diagnosis using a decision tree has been proven to have good performance in terms of classification [18, 19]. The decision tree is a tree-like model that predicts the value of a target variable by learning simple decision rules inferred from the data's features. There are several algorithms for decision tree classification such as ID3 [20], C4.5 [21], CART [22], CHAID [23] and CTree [24]. A typical decision tree consists of a root node, internal nodes and leaf nodes [25] as shown in Figure 2. The arcs from one node to another node denote the conditions of the features. The leaf node represents target variables. Train data are used to build the tree model. Then, the tree model is pruned to check for overfitting and noise. Finally, the optimized tree is used to classify the unlabeled test data.

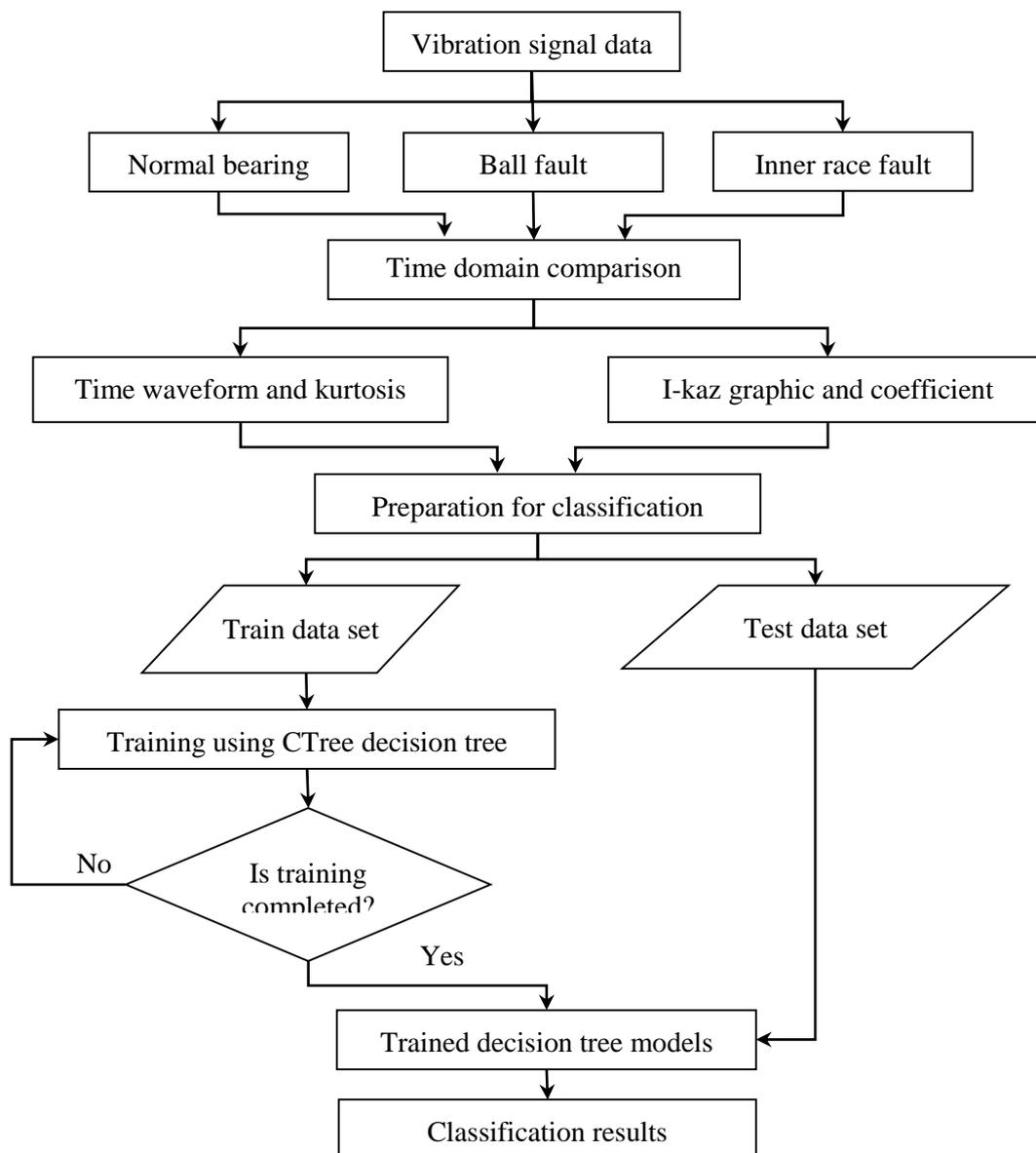


Figure 3. Proposed research process flowchart.

METHODOLOGY

The process flow chart for this research is shown in Figure 3. The vibration signal data are obtained from the Bearing Data Center which is supported by the Case Western Reserve University [26]. This database is widely used by researchers [5, 27-29] especially for testing new diagnostic algorithms with a recognised benchmark study [12].

Experimental Setup

As shown in Figure 4, the test rig for this experiment consists of a 2 HP Reliance induction motor coupled with a torque transducer to measure the torque value applied by a dynamometer via an electronic control system. The test bearings used were the 6205-2RS JEM SKF deep groove ball bearings with dimensions displayed in Table 1. All faulty bearings were artificially damaged using electro-discharge machining (EDM).

Table 1. Test bearing dimensions

Inside Diameter	Outside Diameter	Pitch Diameter	Thickness	Ball Diameter
25 mm	52 mm	39 mm	15 mm	7.94 mm

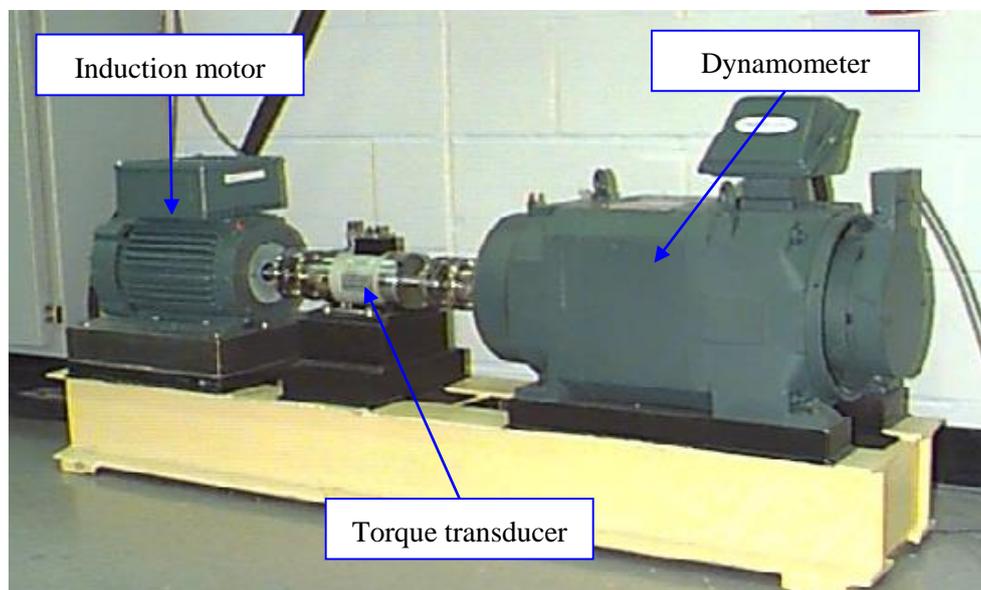


Figure 4. Experimental test rig setup.

Vibration data were collected using accelerometers with a bandwidth of up to 5000 Hz and a 1 V/g accuracy attached using magnetic bases at the 12 o'clock position at the drive end of the motor housing, as Figure 5 shows. A 16 channel DAT recorder was used to record the data. The data were post-processed in the computer using MATLAB software. The experiment was carried out with a very small fault diameter of 0.1778 mm

at a 48 kHz sampling rate and various torques (0, 1, 2 and 3 HP). Basically, a small fault size indicates an early fault stage. Data acquired from the maximum torque (3 HP) were selected for the time-domain analysis. Raw vibration signal data were segregated according to 0.5 seconds segments and features were extracted from each segment with a total of 5 seconds data length.

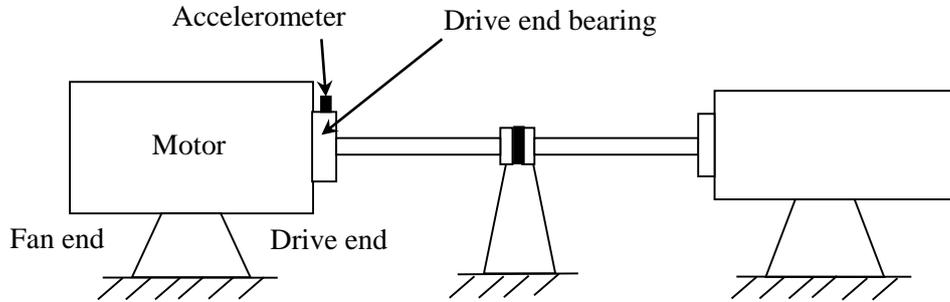


Figure 5. Schematic diagram of experimental setup.

The time waveform together with classic statistical parameters from the first segment were compared for all bearing conditions. The process was repeated with the use of I-kaz graphic and coefficient. Then, the average values of all features were calculated and the percentage differences between the faulty and normal bearings were compared. The I-kaz coefficient is derived from the kurtosis and standard deviation of three frequency ranges, as stated in Eq. (2).

$$Z^{\infty} = \frac{1}{N} \sqrt{Kur_{LF} \cdot s_{LF}^4 + Kur_{HF} \cdot s_{HF}^4 + Kur_{VF} \cdot s_{VF}^4}, \quad (2)$$

where N is sample size; Kur_{LF} , Kur_{HF} and Kur_{VF} are kurtosis; and s_{LF} , s_{HF} and s_{VF} are standard deviation for the low, high and very high-frequency ranges, respectively. All of the frequency ranges were filtered using a Butterworth notch filter where the low-frequency range (LF) is derived from the lowpass of 6 kHz cutoff frequency, the high-frequency range (HF) is from the bandpass of 6 kHz to 12 kHz and the very high-frequency range (VF) is from the highpass of 12 kHz and above.

Decision Tree Classification

The proposed bearing fault diagnosis using decision tree classification in this study started with feature extraction, i.e: kurtosis and I-kaz coefficient from all torque values divided at 0.05 seconds segment intervals. A total of 1200 data were segregated for the training and testing sets at a 70:30 proportion. The decision tree classification is processed using R software [30] with 'caret' package [31]. The training data sets for both features were classified using the CTree algorithm [24] in a separate model using the same control settings as showed in Table 2. Then, all trained models were validated with the testing data set and their accuracies and Kappa values were compared.

Accuracy and Kappa were obtained using equation (3) and (4) respectively.

$$Accuracy = \frac{T}{N} \times 100\%, \quad (3)$$

where T is the number of sample cases correctly classified and N is the total number of sample cases.

$$Kappa = \frac{p_o - p_e}{1 - p_e}, \tag{4}$$

where p_o is the predicted value and p_e is the actual value of classified sample cases.

Table 2. Decision tree training control settings

Description	Value
Method	Cross validation
Fold Number	10
Tune Length	10
Class Probabilities	True

RESULTS AND DISCUSSION

Time-Domain Comparison

The time waveforms with kurtosis values for normal and faulty bearings are shown in Figure 6. Clear differences between the shapes of waveforms for the inner race fault compared to normal bearings are observed in the figure. The time waveform of inner race faulty bearings show a unique repetitive signal pattern, repeated at constant intervals with high amplitude values which matches with the typical pattern reported by Randall and Antoni [32]. However, the ball fault time waveform is similar to those of the normal bearings, making this fault type difficult to be detected. The kurtosis for the normal bearing is close to 3, complying with the Gaussian distribution. The kurtosis value for the inner race faulty bearing is slightly higher than 3 but the ball faulty bearing remains close to 3. Therefore, the time waveform and kurtosis methods are not suitable for the detection of ball faults.

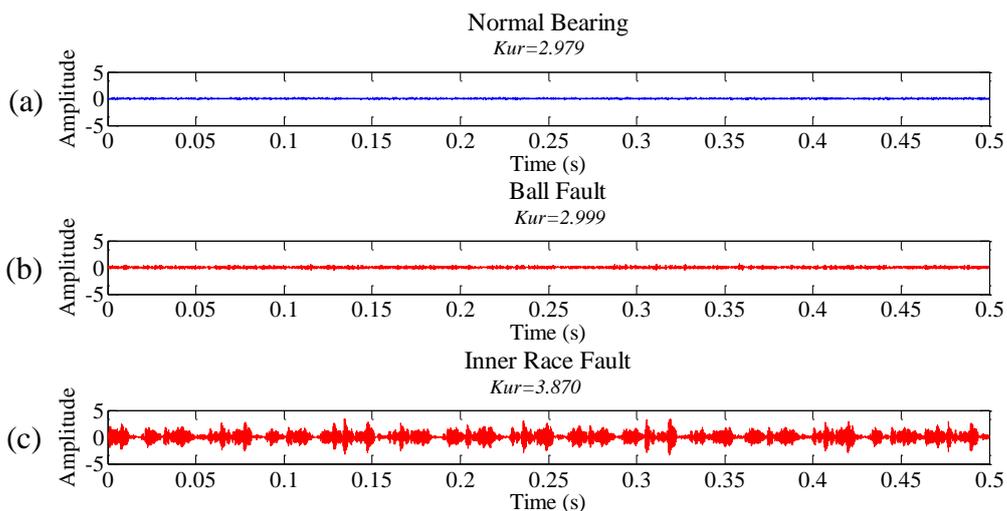


Figure 6. Time waveform with features for (a) normal and (b),(c) faulty bearings

The results for the I-kaz analysis is displayed in Figure 7. The I-kaz graphic for the inner race fault is clear enough to be distinguished from the normal bearings compared to the ball fault. The scatter plot for the inner race fault is bigger than the normal and ball

fault because the data were distributed far off from the mean value. A higher value of I-kaz coefficient for this fault type corroborates with the bigger scatteration of its plots. The inner race fault data for the low-frequency range have the highest amplitude among the two other ranges. Therefore, a big scatteration is observed and the I-kaz coefficient is also higher.

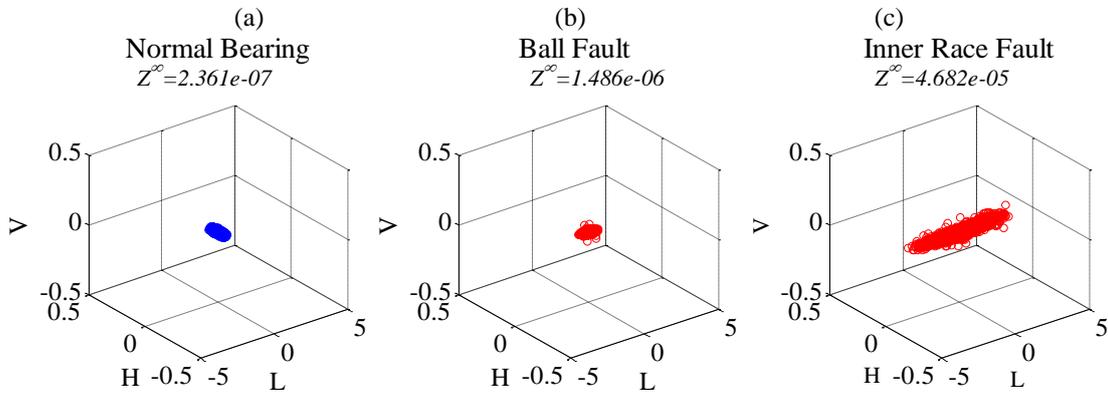


Figure 7. I-kaz graphic and coefficient, Z^∞ for (a) normal and (b),(c) faulty bearings.

A comparison of the percentage difference between fault and normal bearings is tabulated in Table 3 with the kurtosis and I-kaz coefficient average values of 10 samples according to 0.5 second segments. The comparison showed that the I-kaz coefficient was significantly higher than kurtosis for both fault types with 515% and 19693% difference for ball fault and inner race fault respectively compared to kurtosis which had only 4% and 25% difference. The higher percentage difference for I-kaz coefficient is caused by the high value of kurtosis and standard deviation, especially for the low frequency range for inner race faults.

Table 3. Comparison of percentage differences between faulty and normal bearings.

Feature	Normal bearing (baseline value)	Ball fault		Inner race fault	
		Value	% diff.	Value	% diff.
I-kaz coeff. ($\times 10^{-7}$)	2.342	14.402	515%	463.643	19693%
Kurtosis	2.942	3.055	4%	3.666	25%

Decision Tree Classification

The trained decision tree models for the I-kaz coefficient and kurtosis are shown in Figure 8 and the details of the classification results are tabulated in Table 4. The I-kaz coefficient trained model is perfectly distributed from the root node to Node 2 and then branches to an internal node before split to leaf nodes 4 and 5. Node 2, 4 and 5 are classified as normal bearing, ball fault and inner race fault, respectively. The training accuracy for this model is 99.64% and the Kappa coefficient is 0.9946. The trained decision tree model for kurtosis generated 9 nodes which contain three internal nodes before split to five leaf nodes that contain a combination of all bearing conditions. Leaf nodes 4, 5, 6, 8 and 9 contain a total number of 141, 309, 168, 108 and 114 classified targets, respectively. The first three leaf nodes are a mixture of all bearing conditions, whereas the others represent only faulty bearings. The training accuracy for this decision

tree model is significantly lower than the I-kaz coefficient at 63.57%, and the Kappa coefficient is only 0.536.

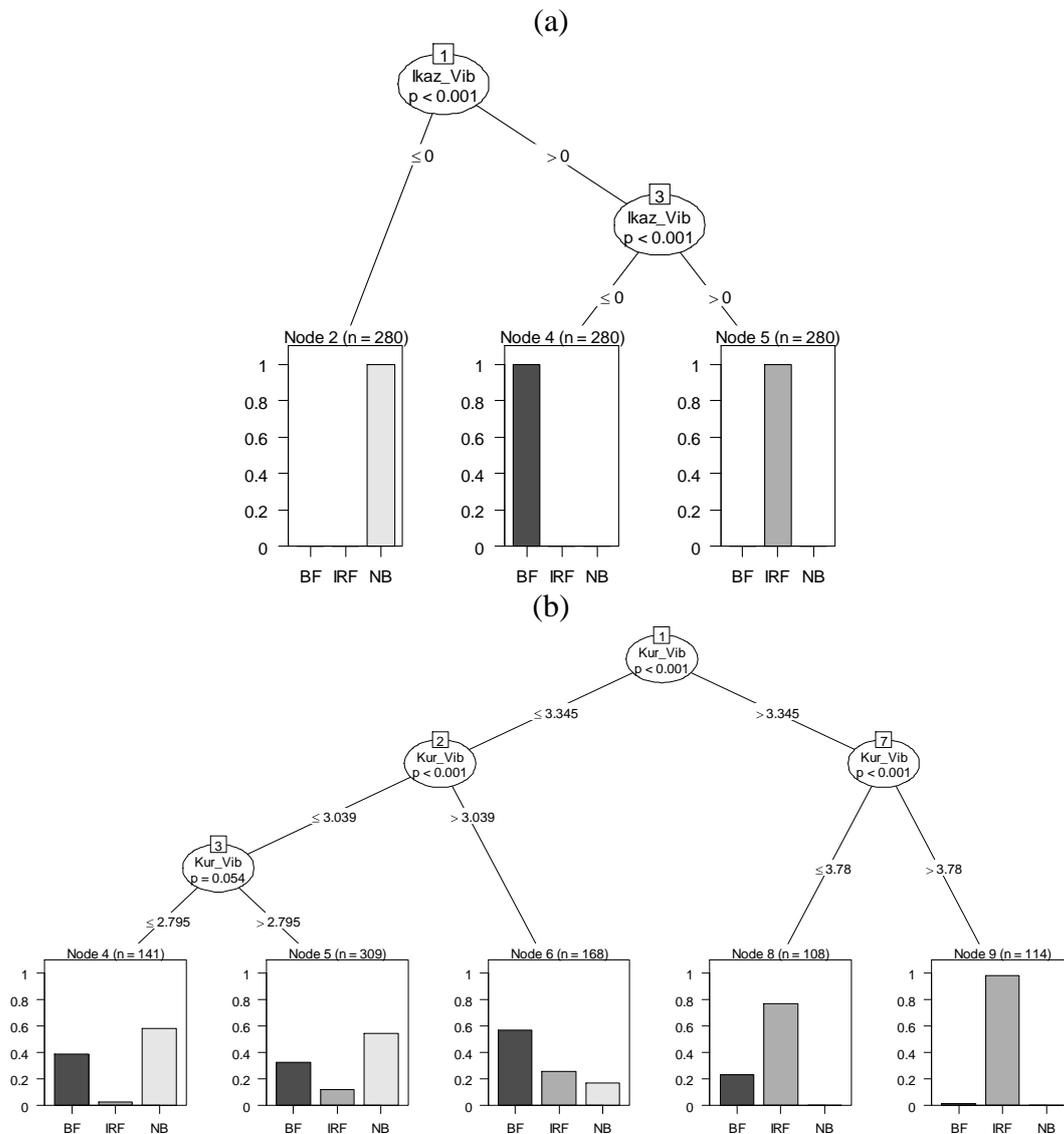


Figure 8. (a) I-kaz coefficient and (b) kurtosis trained decision tree models

Table 4. Comparison of decision tree classification training and testing results using I-kaz coefficient and kurtosis

Feature	Node size	Training		Testing	
		Accuracy	Kappa	Accuracy	Kappa
I-kaz coeff.	5	99.64%	0.9946	100%	1
Kurtosis	9	63.57%	0.4536	65%	0.475

Significant results are observed in the I-kaz coefficient testing classification where the accuracy is 100% and the Kappa coefficient is 1, whereas the kurtosis accuracy is only

65% and the Kappa coefficient is 0.475. The high increment in testing accuracy is possibly caused by the usage of same data for the training and testing sets [29]. The decision tree classification using the I-kaz coefficient was able to detect ball faults which were not diagnosable under any of the applied methods used by the benchmark study [12].

CONCLUSIONS

In this paper, we demonstrate the usage of kurtosis and the I-kaz™ method for the quick detection of bearing faults using raw vibration signal data. In the time-domain, both time waveform and I-kaz graphic displayed clear identification to distinguish the inner race faulty bearings from the normal bearings. Also, the I-kaz and kurtosis values for the faulty bearings had high percentage differences from the normal bearing. However, ball faulty bearings cannot be detected in both plots using visual comparison. Although the I-kaz graphic is inadequate to separate between faulty balls and normal bearings, the I-kaz coefficient provided a significant percentage difference of 515% compared to kurtosis, which is only 4%. The results for the decision tree classification show that the I-kaz coefficient is significantly better with higher training accuracy and Kappa coefficient at 99.64% and 0.9946 respectively compared to kurtosis at only 63.57% and 0.4536. Furthermore, all test data are classified accordingly when using the I-kaz coefficient with 100% accuracy and 1 Kappa value, whereas only 65% is correctly classified at 0.475 Kappa value using the kurtosis test data. It has been proven that the I-kaz analysis is suitable for the diagnostic of bearing faults when using vibration time signals and it can also be applied as a classification feature. This method also performs better than the benchmark study which cannot detect ball faults using any of its applied methods.

ACKNOWLEDGEMENTS

The authors would like to thank the Faculty of Engineering and Built Environment of Universiti Kebangsaan Malaysia for their support during this research work and the Bearing Data Center, Case Western Reserve University for providing ball bearing vibration test data.

REFERENCES

- [1] Karmakar S, Chattopadhyay S, Mitra M, Sengupta S. Induction motor fault diagnosis: Approach through Current Signature Analysis. Singapore: Springer; 2016.
- [2] Toliyat HA, Nandi S, Choi S, Meshgin-Kelk H. Electric machines: Modeling, condition monitoring, and fault diagnosis. Boca Raton: CRC Press; 2016.
- [3] Srihari PV, Govindarajulu K, Ramachandra K. A method to improve reliability of gearbox fault detection with artificial neural networks. International Journal of Automotive and Mechanical Engineering. 2010;2:221-30.
- [4] Lee JY. Sound and vibration signal analysis using improved short-time fourier representation. International Journal of Automotive and Mechanical Engineering. 2013;7:811-9.
- [5] Boudiaf A, Moussaoui A, Dahane A, Atoui I. A comparative study of various methods of bearing faults diagnosis using the Case Western Reserve University data. Journal of Failure Analysis and Prevention. 2016;16:271–84.

- [6] Li W, Wu B, Zhu Z, Qiu M, Zhou G. Fault diagnosis of rolling element bearings with a spectrum searching method. eprint arXiv:151103174. 2015:28 pages.
- [7] Supeni EE, Epaarachchi JA, Islam MM, Lau KT. Development of artificial neural network model in predicting performance of the smart wind turbine blade. *Journal of Mechanical Engineering and Sciences*. 2014;6:734-45.
- [8] Batista L, Badri B, Sabourin R, Thomas M. A classifier fusion system for bearing fault diagnosis. *Expert Systems with Applications*. 2013;40:6788–97.
- [9] Jin X, Zhao M, Chow TWS, Pecht M. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Transactions on Industrial Electronics*. 2014;61:2441–51.
- [10] Jardine AKS, Lin D, Banjevic D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*. 2006;20:1483-510.
- [11] Dyer D, Stewart RM. Detection of rolling element bearing damage by statistical vibration analysis. *Journal of Mechanical Design*. 1978;100:229–35.
- [12] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mechanical Systems and Signal Processing*. 2015;64:100-31.
- [13] Nuawi MZ, Nor MJM, Jamaludin N, Abdullah S, Lamin F, Nizwan CKE. Development of integrated kurtosis-based algorithm for z-filter technique. *Journal of Applied Sciences*. 2008;8:1541-7.
- [14] Mansor NII, Ghazali MJ, Nuawi MZ, Kamal SEM. Monitoring bearing condition using airborne sound. *International Journal of Mechanical and Materials Engineering*. 2009;4:152-5.
- [15] Nuawi MZ, Abdullah S, Ismail AR, Zulkifli R, Zakaria MK, Hussin MFH. A Study on ultrasonic signals processing generated from automobile engine block using statistical analysis. *WSEAS Transactions on Signal Processing*. 2008;5:279-88.
- [16] Abdullah S, Ismail N, Nuawi MZ, Nopiah MZ, Ariffin A. Using the hybrid kurtosis-based method to correlate strain and vibration signals *WSEAS Transactions on Signal Processing*. 2010;3:79-90.
- [17] Ahmad M, Nuawi MZ, Abdullah S, Wahid Z, Karim Z, Dirhamsyah M. Development of tool wear machining monitoring using novel statistical analysis method, I-kaz™. *Procedia Engineering*. 2015;101:355-62.
- [18] Amarnath M, Sugumaran V, Kumar H. Exploiting sound signals for fault diagnosis of bearings using decision tree. *Measurement*. 2013;46:1250–6.
- [19] Nguyen N-T, Kwon J-M, Lee H-H. A study on machine fault diagnosis using decision tree. *Journal of Electrical Engineering and Technology*. 2007;2:461-7.
- [20] Quinlan JR. Induction of decision trees. *Machine Learning*. 1986;1: 81-106.
- [21] Quinlan JR. *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kaufmann Publishers Inc.; 1993.
- [22] Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*: CRC Press; 1998.
- [23] Kass GV. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*. 1980;29: 119-27.
- [24] Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*. 2006;15:651–74.

- [25] Rokach L, Maimon O. Classification Trees. Data Mining and Knowledge Discovery Handbook. 2nd ed: Springer; 2010. p. 149-74.
- [26] Loparo K. Bearing Data Center, Case Western Reserve University - Seeded fault test data. 2003.
- [27] Zhang S, Li W. Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM. Mathematical Problems in Engineering. 2014; 1-10.
- [28] Brkovic A, Gajic D, Gligorijevic J, Savic-Gajic I, Georgieva O, Gennaro SD. Early fault detection and diagnosis in bearings based on logarithmic energy entropy and statistical pattern recognition. 2nd International Electronic Conference on Entropy and Its Applications, 2015. 1- 14.
- [29] Shen C, Wang D, Liu Y, Kong F, Tse PW. Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines. Smart Structures and Systems. 2014;13:453-71.
- [30] Team RC. R: A language and environment for statistical computing. 2015.
- [31] Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. caret: Classification and Regression Training. R package version 6.0-52. 2015.
- [32] Randall RB, Antoni J. Rolling element bearing diagnostics - a tutorial. Mechanical Systems and Signal Processing. 2011;25:485-520.

Nomenclature

caret	Classification and regression training
CART	Classification and regression tree
CHAID	Chi-squared automatic interaction detector
Coeff.	Coefficient
CTree	Conditional inference tree
DAT	Digital audio tape
Diff.	Difference
EDM	Electro-discharge machining HF High-frequency range
HP	Horsepower
ID3	Iterative dichotomized tree
I-kaz	Integrated kurtosis-based algorithm for Z filter
Kur	Kurtosis
LF	Low-frequency range
RMS	Root mean square
VF	Very high-frequency range

Symbols

f_{\max}	Half value of the data sampling rate
N	Data size
μ	Mean
s	Standard deviation
Z^{∞}	I-kaz coefficient