

A histogram statistical method for the detection of localized faults in deep groove ball bearing

Shang-Chih Lin* and Yennun Huang

Research Center for Information Technology Innovation, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, Taiwan

Abstract. This study aims to use the histogram statistical method to establish a deep groove ball bearing fault diagnosis strategy. First, statistical indicators are used to excavate the fault characteristics buried in the vibration signal, and use the histogram to define the characteristic area for fault diagnosis. The results show that the indicators 1, 3, 6 have better statistical differences. Based on this, the accuracy of pattern recognition for all test data is 100 %. Finally, the statistical significance of ball damage was significant, and the results showed high correlation (56~73 %). The correlation between inner race damage model was 49~57 % and healthy model was 52 %. As the inner race damage and health model in the statistical sense, there are some similar, so there is a relatively high correlation. In the future research work, it will be committed to mining more representative indicators to enhance the relevance of abnormal characteristics.

1 Introduction

In the 21st century depends on human technological breakthroughs, gave birth to the history of the fourth industrial revolution. There will be a large number of mechanical equipment towards the development of automation, means that machinery must maintain long-term and uninterrupted operation, in order to maintain active production behavior. However, in the precision machining industry, if the machine can not effectively detect the abnormal situation, long-term will deepen the seriousness of mechanical failure, and finally lead to production line shut down, shortened equipment life and operator safety threats and other issues. In addition, the machine tool is one of the main core facilities of precision machinery, and its closely related industries such as aerospace, shipbuilding, automobile, factory, generator. Of course, there are many factors that affect the normal operation of the machine, such as materials, process, thermal deformation, cutting tools, rotating machinery, processing conditions.

In the course of research on rotating machines, Chang et al. [1-10] used in-depth analysis of rotating machinery, using different techniques [1], such as electrical detection, vibration detection and partial discharge detection. However, the characteristics of the vibration signal are analyzed by the axis trajectory technique, and the characteristic parameters are extracted by using the fractal theory. Finally, the neural network is used to identify the fault [2-3]. In addition, the use of electrical detection method [4], partial discharge detection method [5] to analyze the impact of rotating

machinery failure on the severity of the operating state, and the use of partial discharge detection technology [6-7] diagnosis of high voltage motor stator coil insulation. Further mixing the different signals, such as electrical signals combined with vibration signals [8-9], electrical signals combined with partial discharge signals [10], to develop a more robust expert decision-making system.

However, when the rotating machinery gradually produce abnormal state, often the abnormal current, vibration, temperature, noise and other phenomena, affecting the machine tool quality and personnel safety issues. Antonio et al. [11] used the neural network-like algorithm to predict the surface roughness of the process. Palani et al. [12] applied the machine vision technique to detect the surface condition, and studied the roughness model for the cutting parameters and surface roughness images to find out the relationship between the roughness and the design to ensure that the design of the processing conditions met the quality requirements. Hazza et al. [13] used SVM, EML and ANN and other prediction methods for signal analysis, to find the relationship between the vibration surface roughness, the predicted value and the actual measurement error. Dahbi et al. [14] analyzed the processing conditions, such as cutting speed, feed rate, depth of cut and tool nose radius, so as to learn from the surface roughness of the neural network, from the forecast results can be learned very well, the error range within 10 %. Srinivasa Pai et al. [15] considered the critical effect of tool radius in processing, and then used multi-layer sensing neural network to predict surface roughness. The results show

* Corresponding author: scliciti@citi.sinica.edu.tw

that the predicted value has a roughness parameter close to the actual measurement.

In recent years, Lin et al. [16] discussed the fuzzy diagnosis strategy of rotating electrical machinery in the event of inner ring damage by statistic method. In order to analyze the fault diagnosis technology more deeply, this study aims to use the histogram statistical method to establish a deep groove ball bearing fault diagnosis strategy. Using the statistical indicators to excavate the fault characteristics of the vibration signal, and use the histogram to define the characteristic area of the model, and then achieve the purpose of effective fault diagnosis, detailed method described in Section 2, the Section 3 for the results and analysis, the Section 4 is the conclusion.

2 Methodology

In the analysis method, mainly the use of statistical methods for data analysis, in order to master the deep groove ball bearing state. First of all, the use of statistical indicators to dig out the abnormal characteristics of the data, and by the three-dimensional distribution determine the key indicators, and finally use the histogram statistical method for abnormal diagnosis, detailed process shown in Fig. 1.

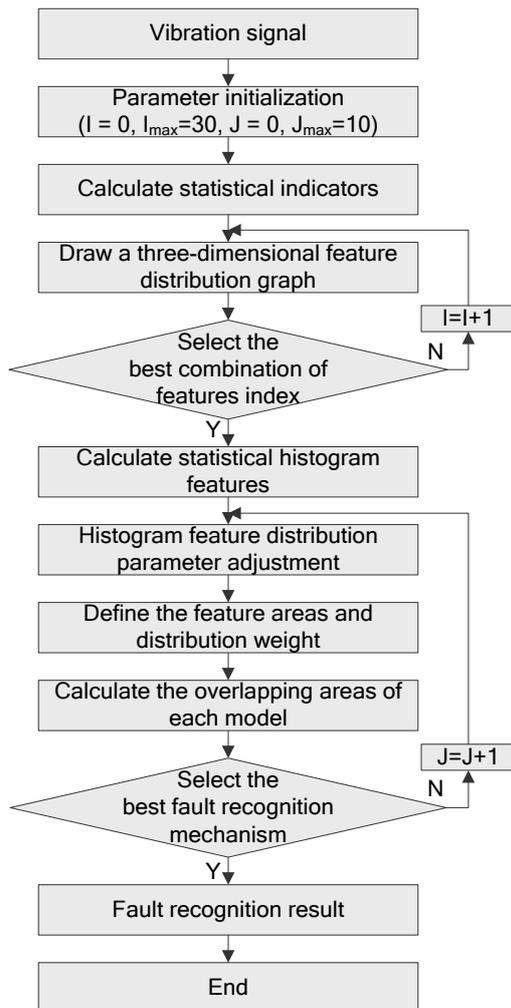


Fig. 1. The proposed method flow chart.

2.1 Feature indicators

In the evaluation of the characteristic index, nine statistical methods were used for the analysis. The main functions used are (1)-(9), such as standard deviation, rms value, peak value, crest factor, kurtosis, skewness, upper bound value of histogram, lower bound value of histogram and mean value.

1) Standard deviation

From the standard deviation index can be viewed vibration intensity distribution in the proportion of the average around, when the values are quite consistent, the standard deviation is relatively small.

$$\text{Standard Deviation} = \sqrt{1/N \sum_{i=1}^N (X_i - \bar{X})^2} \quad (1)$$

where X is the data, i is the number of data, N is the total of data.

2) Root mean square value

From the RMS value indicators, it can be learned that with the mechanical aging, the value will gradually rise, and when the occurrence of mechanical anomalies, the vibration change is more obvious.

$$\text{RMS Value} = \sqrt{1/N \sum_{i=1}^N (X_i)^2} \quad (2)$$

where X is the data, i is the number of data, N is the total of data.

3) Peak value

From the peak value reflects the intensity of the vibration signal, when the abnormal situation occurs, the reaction is often strong vibration intensity.

$$\text{Peak Value} = \max(X_i) \quad (3)$$

where X is the data, i is the number of data.

4) Crest factor

From the crest factor, we can see the ratio of the peak and the RMS value, and observe the change of the vibration waveform.

$$\text{Crest Factor} = \max(|X_i|) / \sqrt{1/N \sum_{i=1}^N (X_i)^2} \quad (4)$$

where X is the data, i is the number of data, N is the total of data.

5) Kurtosis

From the kurtosis index to observe the vibration pulse characteristics, the normal operation showed a gentle Gaussian distribution, and in the object damage, it presents a high peak phenomenon.

$$\text{Kurtosis} = 1/N \sum_{i=1}^N \left(X_i - \bar{X} / \sqrt{1/N \sum_{i=1}^N (X_i - \bar{X})^2} \right)^4 \quad (5)$$

where X is the data, i is the number of data, N is the total of data.

6) Skewness

From the skewness index, observe the value on the left or right side of the mean value, so that the signal is positive or negative skew, but if the value is zero, it means that the values are distributed on both sides of the mean.

$$Skewness = 1/N \sum_{i=1}^N \left(X_i - \bar{X} / \sqrt{1/N \sum_{i=1}^N (X_i)^2} \right)^3 \quad (6)$$

where X is the data, i is the number of data, N is the total of data.

7) Upper bound value of histogram

From the upper bound value of histogram indicators to observe the distribution of the situation, according to the needs of the analysis to dismantle the different distance of data for vibration intensity analysis.

$$UB = \max X_i + 0.5(\max X_i - \min(X_i)/N - 1) \quad (7)$$

where X is the data, i is the number of data, N is the total of data.

8) Lower bound value of histogram

From the upper bound value of histogram indicators to observe the distribution of the situation, according to the needs of the analysis to dismantle the different distance of data for vibration intensity analysis.

$$LB = \min X_i - 0.5(\max X_i - \min(X_i)/N - 1) \quad (8)$$

where X is the data, i is the number of data, N is the total of data.

9) Mean value

From the peak value reflects the intensity of the vibration signal, when the abnormal situation occurs, the reaction is often strong vibration intensity.

$$Mean \ Value = average(X_i) \quad (9)$$

where X is the data, i is the number of data.

When the vibration signal is calculated by the characteristic index, then through the three-dimensional feature distribution to determine the applicability of indicators, and a total of 30 combinations such as 1-2-3, 1-2-4, 1-2-5, 1-2-6, 1-2-7, 1-2-8, 1-2-9, ..., 1-8-9.

2.2 Histogram

In the multiple index analysis, the use of histograms to calculate the characteristics of the distribution area, and two-dimensional graphics were presented statistical samples and measures, as shown in Fig. 2. At the same time, in order to make the statistical samples of each experimental model have significant difference, the fuzzy calculation is carried out for the characteristic area, so as to improve the performance of fault identification.

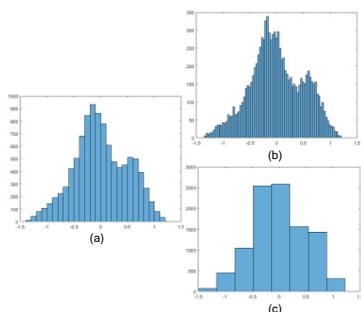


Fig. 2. Histogram statistical distribution: (a) original bins, (b) more bins and (c) fewer bins.

Then use (10) to calculate the weight of each feature distribution area, to infer the possible fault type.

$$A(x)_{w_y} = y_i / (i^{\max} \cdot z^{\max}) \quad (10)$$

where A is the experimental model, x is the category, the range is 1, 2, 3, respectively, health, inner race damage and ball damage model. W is the weight, y is the characteristic distribution area, the range is 1, 2, ..., f , f is the fuzzy parameter. i is the amount of data, the number is 12. z is the best indicator, the number is 3.

The third section will detail the characteristics of the experimental model and its fault recognition performance.

3 Result and Analysis

In the experimental analysis section, the first use of data (number of twelve) to establish a statistical model, and the use of data (number of nine) for performance testing. The following will be discussed in two stages, namely, the selection of the best characteristic index (Section 3.1) and fault diagnosis (Section 3.2).

3.1 The select of the best feature index

This section is the characteristic index of the analysis experimental data, and the average is the criterion of the evaluation index. The health model and the ball damage model have similarities in the indicators 4, 5 and 8; while the health model and the inner race damage model have similarities in the index 8; the inner race damage model and the ball damage model are similar in the indicators 8 and 9, the detailed information is shown in Table 1.

Table 1. The statistical relationship between the index and the model.

Index	Fault type		
	H	I	B
1	0.0633	0.2996	0.1471
2	0.0644	0.2996	0.1472
3	0.2423	1.5276	0.5415
4	3.7611	5.0992	3.6788
5	2.9342	5.5637	2.8305
6	-0.1838	0.0901	0.0265
7	0.2423	1.5276	0.5416
8	0.0000	0.0000	0.0000
9	0.0118	0.0046	0.0046

After analysis and comparison, the selection of the most significant differences in the indicators, respectively, indicators 1, 3 and 6. In addition, the three-

dimensional spatial distribution characteristics shown in Figs. 3 and 4. The diagnostic analysis results are described in Section 3.2.

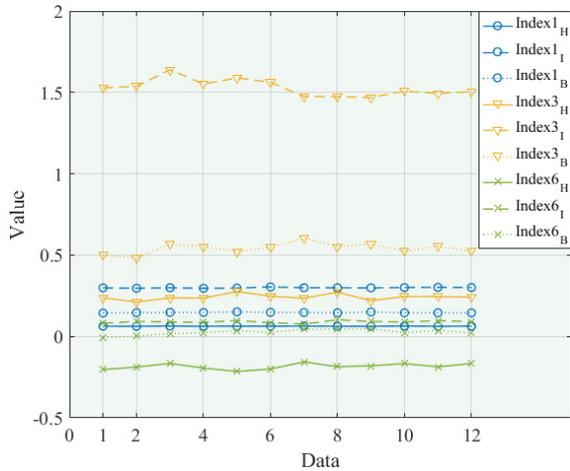


Fig. 3. Comparison of the best feature index of each model.

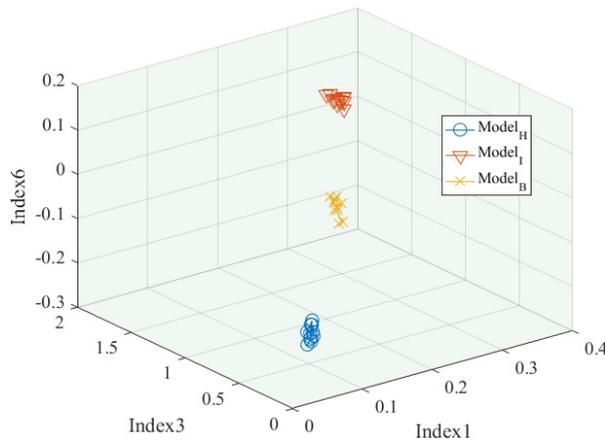


Fig. 4. Three-dimensional characteristic distribution of each experimental model.

3.2 Fault diagnosis

In the fault diagnosis section, the decision analysis is based on the characteristic parameters of Section 3.1.

However, in Section 3.2.1, a histogram feature distribution is used to determine an ideal feature model.

Then compare the differences between the data and the various feature models in Sections 3.2.2 to 3.2.4; Finally, we discuss the relevance of each model in Section 3.2.5 to verify the validity of the proposed method.

3.2.1 The select of the best histogram feature distribution

This section is intended to explore the use of histograms to analyze the performance of statistical indicators and to define valid feature models from them. Mainly it is focused on the selection of the histogram of the feature distance, in order to obtain the greatest difference between the various experimental models. However, in order to obtain the ideal feature model, the bin value of

the histogram with different degrees of fuzz is tested, which can effectively define the characteristic interval of the statistical index. And when the bin value is more, that has a more feature interval.

At the same time, there is a direct and positive effect on the difference of the model, but it will increase the load of the computer; on the other hand, fewer bin values are less accurate for the interpretation of feature metrics and reduce the differences between the various feature models, but the computational load is lighter. In this case, the performance of different bin values is tested, as shown in Fig. 5.

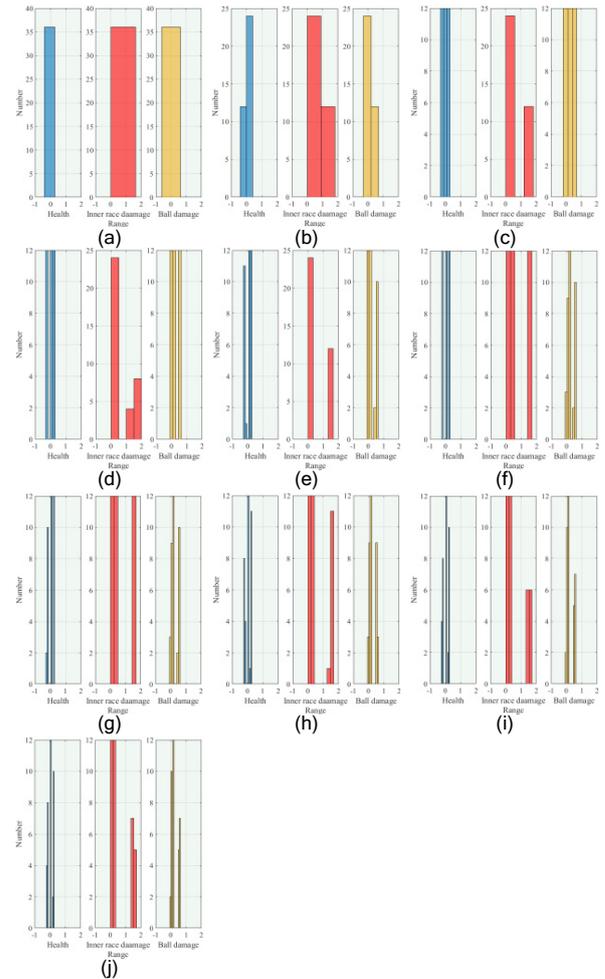


Fig. 5. Histogram feature distribution parameter adjustment: (a) first to (j) tenth.

It is clear from Fig. 5 that the difference in each feature interval is significant. When the bin value is small, the feature area distribution is highly overlapped, which means that each model has high similarity, as shown in Fig. 5 (a); however, if the bin value is more, the feature area is divided more detailed, as shown in Fig. 5 (j).

In order to determine the characteristic interval that is more suitable for each model, we first calculate the overlapping regions of the model with the intersection characteristics (gray in Table 2), then the minimum overlap range as the parameters of each experimental model, as shown in Table 2.

Table 2. Simplified feature model and overlapping area distribution.

No.	Feature area			Overlapping area
	Health	Inner race damage	Ball damage	
1	[-0.40 0.28]	[0.00 1.64]	[-0.60 0.61]	[-0.40 0.61]
2	[-0.40 0.40]	[0.00 1.80]	[-0.30 0.70]	[-0.30 0.70]
3	[-0.30 0.30]	[0.00 0.60] [1.20 1.80]	[-0.20 0.70]	[-0.20 0.60]
4	[-0.30 -0.15] [0.00 0.30]	[0.00 0.50] [1.00 2.00]	[-0.10 0.26] [0.44 0.62]	[0.00 0.50]
5	[-0.27 -0.05] [0.06 0.28]	[0.00 0.33] [1.32 1.65]	[-0.10 0.20] [0.35 0.65]	[-0.10 -0.05] [0.00 0.28]
6	[-0.24 -0.15] [0.03 0.12] [0.21 0.30]	[0.00 0.56] [1.40 1.68]	[-0.10 0.26] [0.38 0.62]	[0.00 0.56]
7	[-0.28 -0.12] [0.04 0.12] [0.20 0.28]	[0.00 0.48] [1.44 1.68]	[-0.08 0.22] [0.42 0.62]	[0.00 0.48]
8	[-0.24 -0.11] [0.02 0.09] [0.15 0.28]	[0.00 0.42] [1.26 1.68]	[-0.07 0.20] [0.47 0.65]	[0.00 0.28]
9	[-0.25 -0.13] [0.05 0.10] [0.16 0.28]	[0.00 0.38] [1.33 1.71]	[-0.06 0.16] [0.46 0.61]	[0.00 0.28]
10	[-0.24 -0.14] [0.02 0.07] [0.18 0.28]	[0.00 0.34] [1.36 1.70]	[-0.06 0.07] [0.14 0.21] [0.48 0.61]	[0.00 0.28]

It is known from Table 2 that each model has a high degree of overlap in the feature areas of numbers 1 and 2.

In addition, to further consider the impact of the distribution of weight, the health model in the number 10 features and weight distribution is most clear.

However, from the sixth fuzzy adjustment to obtain a stable result, and in order to reduce the complexity of modeling and computing resources, so it is the basis for failure decision; and so on, determine the characteristics of the inner race damage and ball damage model, respectively, for the number 10 and 6.

3.2.2 Diagnostic results based on health characteristics

In the health model section, the feature distribution regions are defined by histograms, [-0.24 -0.15], [0.03 0.12] and [0.21 0.30], respectively.

After the data were analyzed by statistical index and histogram, the health model test data (H1, H4, H7) fully accorded with the characteristic area of the health model; the inner race damage model test data (I2, I5, I8) have only a few of them that meet the health characteristics; ball damage model test data (B3, B6, B9) are almost non-compliant with health characteristics, as shown in Table 3.

Table 3. Comparison of test data with health model.

Test data	Feature area & Distribution weight		
	[-0.24 -0.15]	[0.03 0.12]	[0.21 0.30]
	0.33	0.33	0.33
H1	-0.20	0.06	0.24
I2	-	0.08	0.30
B3	-	-	-
H4	-0.17	0.06	0.24
I5	-	0.09	0.30
B6	-	-	-
H7	-0.20	0.06	0.25
I8	-	0.09	-
B9	-	0.03	-

According to the results in Table 3, calculate the weight of the test data. In the diagnostic results, the test data were considered similar to the health status of 52 % (health test data), 24~37 % (inner race damage test data), 0~22 % (ball damage test data), as shown in Fig. 6.

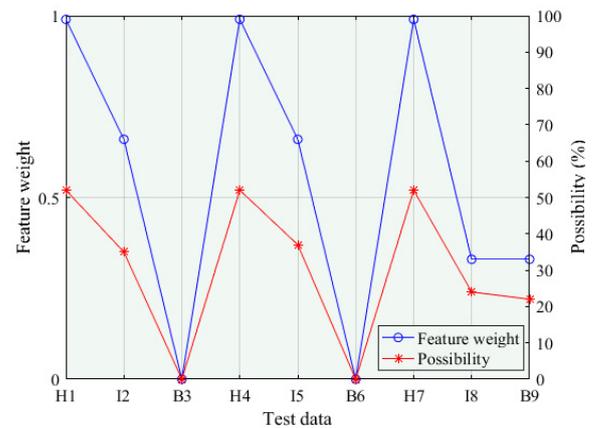


Fig. 6. The similarity coefficients and diagnostic results in healthy feature model.

3.2.3 Diagnostic results based on inner race damage characteristics

In the inner race damage model section, the feature distribution regions are defined by histograms, [0.00 0.17], [0.17 0.34], [1.36 1.53] and [1.53 1.70], respectively.

After the data were analyzed by statistical index and histogram, the health model test data (H1, H4, H7) have only a few of them that meet the inner race damage characteristics; the inner race damage model test data (I2, I5, I8) are very consistent with the inner race damage characteristics; ball damage model test data (B3, B6, B9) are almost non-compliant with inner race damage characteristics, as shown in Table 4.

Table 4. Comparison of test data with inner-race damage model.

Test data	Feature area & Distribution weight			
	[0.00 0.17]	[0.17 0.34]	[1.36 1.53]	[1.53 1.70]
	0.33	0.33	0.22	0.11
H1	0.06	0.24	-	-
I2	0.08	0.30	1.53	1.53
B3	0.14	-	-	-
H4	0.06	0.24	-	-
I5	0.09	0.30	1.51	-
B6	0.02, 0.15	-	-	-
H7	0.06	0.25	-	-
I8	0.09	0.31	-	1.56
B9	0.03, 0.15	-	-	-

According to the results in Table 4, calculate the weight of the test data. In the diagnostic results, the test data were considered similar to the inner race damage status of 35 % (health test data), 49~57 % (inner race damage test data), 22~33 % (ball damage test data), as shown in Fig. 7.

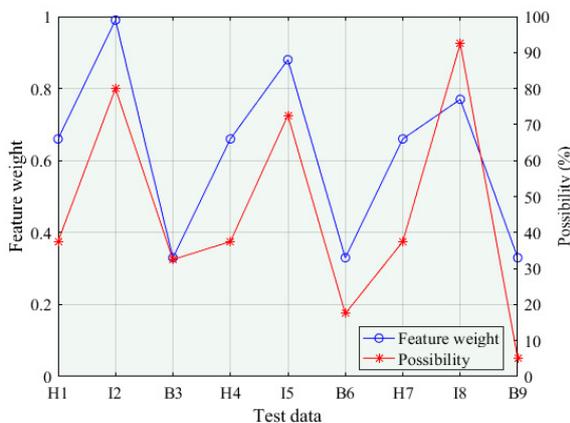


Fig. 7. The similarity coefficients and diagnostic results in inner race damage feature model.

3.2.4 Diagnostic results based on ball damage characteristics

In the ball damage model section, the feature distribution regions are defined by histograms, [-0.07 0.02], [0.02 0.11], [0.11 0.20], [0.47 0.56] and [0.56 0.65], respectively.

Then, the results of the analysis show that the health model test data (H1, H4, H7) and the inner ring damage model test data (I2, I5, I8) do not meet the ball damage characteristics; ball damage model test data (B3, B6, B9) conforms to the characteristic area properties, as shown in Table 5.

Table 5. Comparison of test data with ball damage model.

Test data	Feature area & Distribution weight				
	[-0.07 0.02]	[0.02 0.11]	[0.11 0.20]	[0.47 0.56]	[0.56 0.65]
	0.08	0.25	0.33	0.25	0.08
H1	-	0.06	-	-	-
I2	-	0.08	-	-	-
B3	-0.01	-	0.14	0.50	-
H4	-	0.06	-	-	-
I5	-	0.09	-	-	-
B6	0.02	0.02	0.15	0.52	-
H7	-	0.06	-	-	-
I8	-	0.09	-	-	-
B9	-	0.03	0.15	0.55	-

According to the results in Table 5, calculate the weight of the test data. In the diagnostic results, the test data were considered similar to the ball damage status of 13 % (health test data), 13~19 % (inner race damage test data), 56~73 % (ball damage test data), as shown in Fig. 8.

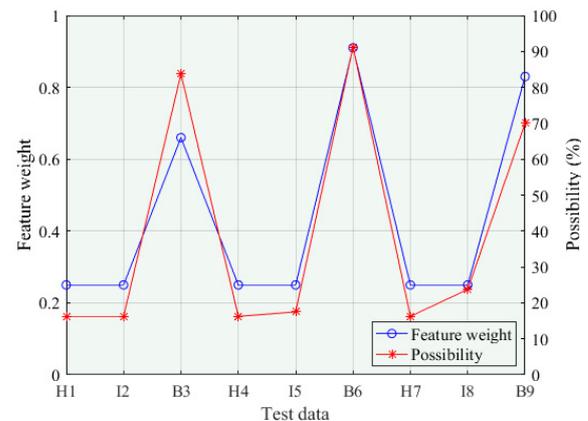


Fig. 8. The similarity coefficients and diagnostic results in ball damage feature model.

3.2.5 Evaluate pattern recognition performance

This section will be based on the analysis of the previous three sections to assess its performance, through the analysis results that all the test data can get the correct pattern recognition results, in addition to prove that the proposed method is feasible, as shown in Fig. 9.

In Fig. 9, it is known that the health model has some common characteristics with the inner ring damage model, but does not produce a misjudgment, and does not have the characteristics of ball damage; however, in the part of the ball damage model, the correct pattern can be clearly determined because it has distinct characteristics from other models, as shown in Fig. 9.

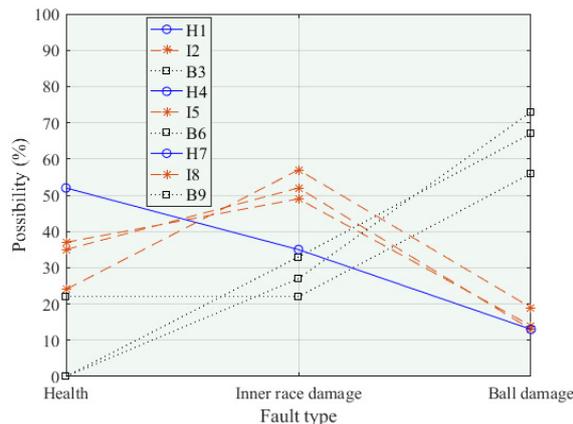


Fig. 9. The deep groove ball bearing defect diagnosis results.

4 Conclusion

In this paper, the statistical features are used to excavate the hidden faults in the vibration signal, and the most suitable indicators are screened for histogram statistics. Finally, the feature regions of each model are used as the basis for fault diagnosis. The results showed that the indexes 1, 3 and 6 had significant differences in characteristics. However, all test data have 100 % fault resolution accuracy. On the other hand, due to the significant difference in the statistical index of the ball damage model, the evaluation results showed high correlation (56~73 %); while the correlation of the inner race damage model was 49~57 %; the health model correlation was 52 %. As the inner race damage and health model in the statistical sense, there are some similar, so there is a relatively high correlation. In the future study will continue to carry out data mining, to find more significant differences in the characteristics of indicators to strengthen the performance of correlation analysis. We sincerely hope that the research results will have more help to the maintenance staff.

Acknowledgements

The work is supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 103-2221-E-001-028-MY3.

References

1. H. C. Chang, C. Cheng, Y. T. Lin, J. F. Chen, C. C. Kuo, H. C. Chen, Y. H. Lin, Y. M. Jheng, S. C. Lin, C. F. Hsieh, and C. Y. Lin, "Smart Grid Development of On-line Monitoring and Diagnosis Methods for High Voltage Motor," *Monthly Journal of Taipower's Engineering*, **822**, pp. 94-115, (2017)
2. H. C. Chang, S. C. Lin, C. C. Kuo, C. Y. Lin, and C. F. Hsieh, "Using Neural Network based on the Shaft Orbit Feature for Online Rotating Machinery Fault Diagnosis," *IEEE International Conference on System Science and Engineering*, pp. 1-4, (2016)
3. H. C. Chang, C. C. Kuo, S. C. Lin, and C. Y. Lin, "Application of fractal theory to the rotating

machinery of the vibration orbit feature extraction," *The 37th Symposium on Electrical Power Engineering*, pp. 1-6, (2016)

4. H. C. Chang, S. C. Lin, C. C. Kuo, and C. F. Hsieh, "Induction Motor Diagnostic System Based on Electrical Detection Method and Fuzzy Algorithm," *International Journal of Fuzzy Systems*, **18**, 5, pp. 732-740, (2016)
5. H. C. Chang, S. C. Lin, C. C. Kuo, and C. Y. Lin, "Fuzzy Theory-Based Partial Discharge Technique for Operating State Diagnosis of High-Voltage Motor," *International Journal of Fuzzy Systems*, **18**, 6, pp. 1092-1103, (2016)
6. H. C. Chang, S. C. Lin, C. C. Kuo, and C. Y. Lin, "Application of Artificial Neural Network to the Partial Discharge Recognition of High-Voltage Motor," *Proceedings of 2016 National Symposium on System Science and Engineering*, pp. 1-4, (2016)
7. H. C. Chang, C. C. Kuo, S. C. Lin, C. Y. Lin, and B. R. Chen, "Partial Discharge Pattern Recognition on the Stator Winding Insulation of High Voltage Motors Using Artificial Neural Network," *The 37th Symposium on Electrical Power Engineering*, pp. 1-6, (2016)
8. H. C. Chang, Y. T. Lin, C. Cheng, C. C. Kuo, S. C. Lin, Y. M. Jheng, C. F. Hsieh, and C. Y. Lin, "Fuzzy-based Fault Diagnosis System for Induction Motor," *The 37th Symposium on Electrical Power Engineering*, pp. 1-6, (2016)
9. H. C. Chang, S. C. Lin, C. C. Kuo, C. F. Hsieh, and C. Y. Lin, "Fuzzy Theory-Based Hybrid Decision-Making System for Rotating Electrical Machinery Fault Diagnosis," *International Conference on Fuzzy Theory and Its Applications*, pp. 1-6, (2016)
10. H. C. Chang, S. C. Lin, C. C. Kuo, C. F. Hsieh, and C. Y. Lin, "A Two-Stage Rotating Machinery Condition Monitoring and Diagnostic System," *International Conference on Engineering and Natural Science-Summer Session*, pp. 136-154, (2016)
11. A. J. Vallejo, M. M. Ruben, R. M. Ricardo, and G. C. Luis, "Online Prediction of Surface Roughness in Peripheral Milling Processes," *European Control Conference*, pp. 3695-3700, (2009)
12. S. Palani and Y. Kesavanarayana, "Prediction of surfaceroughness in end milling process by machine vision using neuro fuzzy network," *International Conference on Science Engineering and Management Research*, pp. 1-5, (2014)
13. M. H. F. A. Hazza, E. Y. T. Adesta, and M. F. Amin, "Seder Using Soft Computing Methods as an Effective Tool in Predicting Surface Roughness," *International Conference on Advanced Computer Science Applications and Technologies*, pp. 9-13, (2015)
14. S. Dahbi, L. Ezzine, and H. E. Moussami, "Modeling of Surface Roughness in Turning Process by using Artificial Neural Networks," *International*

- Conference on Logistics Operations Management*, pp. 1-6, (2016)
15. N. Fang, P. S. Pai, and N. Edwards, "Multidimensional Signal Processing and Modeling with Neural Networks in Metal Machining_Cutting Forces, Vibrations, and Surface Roughness," *IEEE International Conference on Communication Software and Networks*, pp. 77-80, (2016)
 16. S. C. Lin and Y. N. Huang, "Application of Statistical Methods to Excavate the Abnormal State Characteristics of Rotating Machinery," *Proceedings of 2017 National Symposium on System Science and Engineering*, pp. 112-116, (2017)