Development of a Machine Learning Algorithm to Improve Defect Detection Accuracy for Rolling Bearings

Masashi KITAI* Yoshinobu AKAMATSU**
Ken-ichi FUKUI***

Detection of rolling bearing defects is important for machine maintenance. Previous studies by the authors attempted to detect artificial defects of various sizes on the outer ring raceway surface of a rolling bearing, and found that key features for successful detection depend on the defect size. This study shows an improvement in accuracy of the defect detection process by using a size-based feature selection approach coupled with a two-step outlier detection method.

1. Introduction

Rolling bearings are indispensable elements for many rotating machines and their use spans across many industries such as automotive, aircraft, and industrial plants. When rolling bearings suffer damage, it may not only affect the accuracy and operating efficiency of rotating machines but may also have serious consequences for the machines themselves if the damage propagates and expands. Therefore, it is important to accurately detect any damage in rolling bearings.

A frequently used diagnostic approach for rolling bearings is analysis on vibration acceleration or vibration data of Acoustic Emission (AE). The reason for this is because measurement of the data is easy and can be conducted while the target machines are in operation. Methods of detecting damage on rolling bearings using vibration data include, for example, condition monitoring by analyzing changes in trends of different statistical data such as effective value and peakedness calculated from vibration acceleration[1], condition monitoring by tracking changes of characteristic frequency peaks after FFT process[2] and detecting formation of initial cracks using AE[3]. However, these methods require an understanding of vibration characteristics of rolling bearings and signal processing for analysis. Therefore, a simpler diagnostic approach is desirable.

On the other hand, an approach using machine learning has recently been attracting attention as a method to evaluate damage conditions of rotating machines as it does not require an understanding of analysis and signal processing. We can see some examples in faulty symptom detection of hydraulic power plants[5] using the One Class Support Vector Machine (OC SVM)[4], faulty vibration detection of driving machines[7] using Nearest Neighbor Data Description[6] and fault detection of wind turbines[9] using the tandem connection approach of the Deep Neural Network/Gaussian Mixture Model[8].

We also attempted fault detection by OC SVM, creating artificial faults of different sizes on outer raceways of rolling bearings, and found that small defects could not be detected depending on the rotational speed of the main shaft[10]. After examining various feature quantities (considering domain, frequency filter, and measurement direction of vibration acceleration), we verified that variations in defect size change feature quantities that are useful for defect detection[11]. Therefore, it is necessary to conduct feature selection against different defect sizes to improve the accuracy of defect detection.

In this study, we propose a new defect detection method combining feature selection against various artificial defect sizes and the two-step outlier detection method. As a preliminary test, we selected Local Outlier Factor (LOF)[12] as a method to achieve the highest defect detection accuracy, after comparing the defect detection accuracy of various outlier detection methods. Additionally, as a result of investigation into classification accuracy of normal bearings and bearings with small defects by Random Forest (RF)[13], it was determined that RF was appropriate to capture features of vibration due to defects as it could achieve high classification accuracy. Therefore, RF was adopted for feature selection against bearings with artificial defects of different sizes and normal bearings, after which the effect of feature selection by LOF on defect detection accuracy was examined.

Finally, the proposed method considering preliminary test results and the original method were compared to evaluate the defect detection performance of the proposed method. By conducting feature selection according to defect sizes and two-

* Product and Business Strategic Planning Dept., New Product and Business Strategic Planning Headquarters
** New Product and Business Strategic Planning Headquarters
*** The Institute of Scientific and Industrial Research, Osaka University

step outlier detection, it was shown that the defect detection accuracy was improved compared to the existing research results[10) 11)].

In the following section, relevant research regarding fault detection of rolling bearings using machine learning will be reviewed. In Section 3, the outlier detection method and feature selection method that are used in this research are discussed. In Section 4, the evaluation test and results of the proposed method are presented along with the final conclusion.

2. Relevant research

Recently, diagnostic approaches using machine learning and a combination of machine learning and vibration analysis have been attracting attention as diagnostic approaches for rolling bearings and rotating machines that involve rolling bearings.

For example, Li et al. succeeded in determining a fault location for rolling bearings with higher accuracy than the existing methods, targeting rolling bearings with defects set on surfaces of the inner and outer ring raceways and rolling elements[18)], by feature extraction with Local Mean Decomposition[14)] and Multiscale Permutation Entropy[15)], feature selection with Laplacian Score (LS)[16)], and finally by classification with Improved SVM-BT, which is their own unique improvement of the Support Vector Machine based of Binary Tree (SVM-BT)[17)]. Also, Bugharbee et al. improved detection accuracy of faults on surfaces of inner and outer ring raceways and rolling elements of bearings under different rotational speeds by analysis of main components using Singular Spectrum Analysis[19)] and the creation of a fault detection threshold based on Mahalanobis' Distance, using vibration data of normal bearings[20)]. Shao et al. classified defects using Deep Belief Network[22)] which optimizes the number of neurons in each hidden layer by Particle Swarm Optimization[21)] targeting rolling bearings with defects of different sizes on the outer and inner ring raceways and rolling elements. Additionally, they showed its advantage of classification accuracy over other classification methods such as SVM and Bayes' estimation[23)].

However, Bugharbee et al. target bearings with only one type of defect in each part of the outer and inner raceway and rolling element surfaces and do not state the impact of the size of defects on detection accuracy[20)]. Li and Shao classify defects by supervised learning as the method of defect detection and do not discuss defect detection accuracy under unsupervised learning[18) 23)].

3. Proposed method

3.1 Overview

The proposed method takes the input data v, which is the feature vector calculated from measured vibration acceleration data of rolling bearings, and calculates the anomaly score a by feature selection using the classification method and two-step outlier detection method. **Fig. 1** shows a flow diagram of defect detection using the proposed method. The defects to be detected are divided into 3 stages based on their status. The importance of each feature is calculated using the classification method for each defect status, after which the feature selection is performed (Feature Selection in **Fig. 1**).

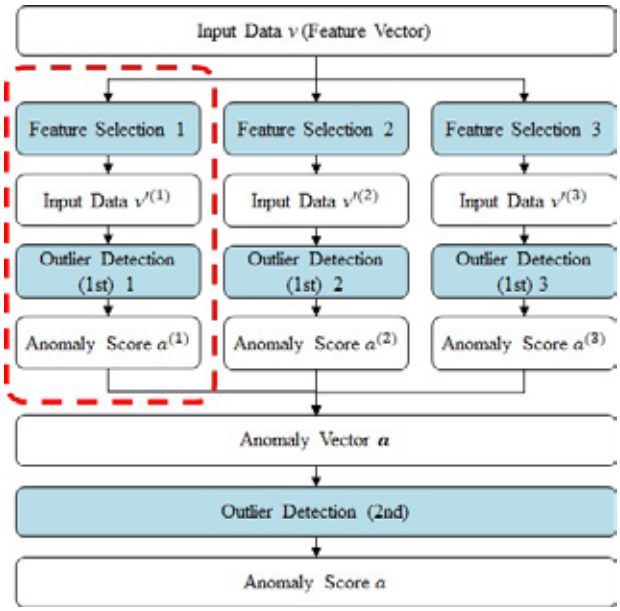


Fig. 1 Flow of defect detection

The calculation of importance by the classification method is only done during training. During testing, feature selection is directly performed using the importance calculated during training. After feature selection, each outlier detection method is trained and tested individually for each input data $\boldsymbol{v}'^{(M)}$, $M \in \{1\ 2\ 3\}$ and the respective anomaly score $a^{(M)}$ is calculated (Outlier Detection (1st) 1 to Outlier Detection (1st) 3 in **Fig. 1**). The input data after feature selection $\boldsymbol{v}'^{(M)}$ are targeting different defect statuses, respectively, and therefore defect statuses other than their target may not be detectable. Therefore, multiple anomaly scores $a^{(M)}$ obtained from input data $\boldsymbol{v}'^{(M)}$ are handled as one vector (anomaly vector $\boldsymbol{a}$ in **Fig. 1**) for corresponding measured data. A final anomaly score a is calculated by outlier detection against the anomaly vector $\boldsymbol{a}$ (Outlier Detection (2nd) in **Fig. 1**). In this research, after the anomaly score is calculated based on the proposed method on all the input data obtained from the measured data of the target to be tested. The ratio of input data for which the anomaly score exceeds the defect detection threshold determined in advance is evaluated as an anomaly

ratio. The presence of any defect in the target data is determined based on the magnitude of the anomaly ratio.

3.2 Input/output data

The input data $\boldsymbol{v}$ are the vectorized statistics of the measured vibration acceleration data divided into segments at a fixed time interval considering domain (time, frequency, and quefrency), band pass filter (BPF), and the direction of sensor measurement. The output data is the anomaly score a obtained by the outlier detection method of the second step.

3.3 Evaluation metrics

Anomaly Ratio (AR) and Area Under Curve Score (AUC Score)[24] are used as metrics to evaluate defect detection accuracy using the proposed method. The calculation method of AR and AUC Score is shown in the following:

3.3.1 Anomaly ratio

The anomaly ratio AR is calculated using the following equation based on anomaly score a which is calculated by the proposed method against the input data $\boldsymbol{v}$ that can be obtained from consecutive multiple segments in the measurement data. $a_{\text{threshold}}$ is a defect detection threshold to be predetermined and set as the average of the anomaly score against the normal data + 5 × standard deviation of the anomaly score against the normal data. In addition, N_{DS} indicates the number of segments included in the measurement data of the evaluation target.

$$U_i = \begin{cases} 1(a_i \geq a_{\text{threshold}}) \\ 0(a_i \leq a_{\text{threshold}}) \end{cases} \quad (1)$$

$$AR = \frac{1}{|N_{DS}|} \sum_{i \in N_{DS}} U_i \quad (2)$$

3.3.2 Area Under Curve Score

As evaluation metrics of identification performance, when positive is correctly identified as positive, it is called True Positive (TP); when negative is correctly identified as negative, it is called True Negative (TN); when negative is incorrectly identified as positive, it is called False Positive (FP); and when positive is incorrectly identified as negative, it is called False Negative (FN). The False Positive Rate (FPR) and True Positive Rate (TPR) are given in the following equations, respectively:

$$FPR = \frac{FP}{FP + TN} \quad (3)$$

$$TPR = \frac{TP}{TP + FN} \quad (4)$$

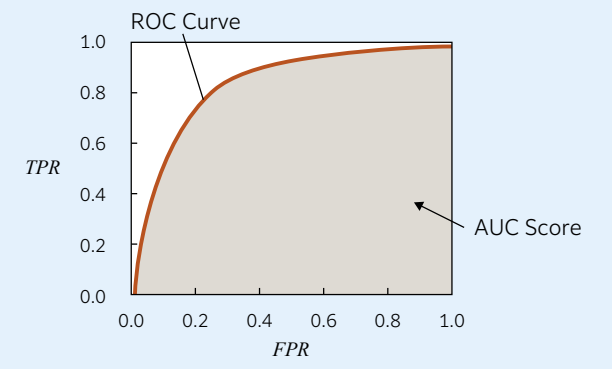


Fig. 2 ROC Curve and AUC Score

By setting FPR at a certain defect detection threshold on the horizontal axis and TPR on the vertical axis and plotting the relation of FPR and TPR in the graph when the defect detection threshold is changed, the Receiver Operating Characteristic Curve (ROC Curve) is obtained. An example of ROC Curve is shown in **Fig. 2**. AUC Score is defined as the area below the ROC Curve. If the normal data and anomaly data can be completely separated, the AUC Score becomes 1.0. On the other hand, in random identification, the AUC score will be 0.5. This means that, if the AUC Score is close to 1.0, then the created machine learning model has a high identification performance.

3.4 Machine learning method

For the outlier detection method, OC SVM, LOF, and Isolation Forest (IF)[25] were used and for the feature selection method, RF and LS were used. The following is a discussion of each of these methods. For implementation of the machine language method, Python 2.7 and Scikit-learn 0.19.0 were used.

3.4.1 One Class Support Vector Machine

OC SVM is a method to determine the identification boundary that classifies normal data and anomaly data by mapping normal data to a certain feature space and obtaining the radius and center of the hypersphere where normal data is preferably included in the feature space. When unknown data is obtained and mapped to the feature space, if it is mapped outside the radius of the hypersphere, the target data is regarded as an anomaly. By selecting the kernel function used for mapping into feature space, non-linear problems can be handled. In this research, Gaussian kernel was used for the kernel function.

3.4.2 Local Outlier Factor

LOF is an outlier detection method based on the density of the feature space. Outliers are based on an assumption that they distribute in a range with low density in the feature space. For each point on the feature space, the marginal density is calculated. When the marginal density of the target point and neighboring points are almost the same, the target point is regarded as normal. On the other hand, when

the marginal density of the target point is lower than the marginal density of the neighboring points, the target point is regarded as an anomaly.

3.4.3 Isolation Forest

IF is a method to calculate the anomaly score based on the average number of splits until certain data is isolated from the other data by random selection of feature quantities and split points. Since normal data have similar characteristics to other normal data, the average number of splits for each normal data to become isolated will be large. On the other hand, since anomaly data have different characteristics from normal data, the average number of splits for each normal data to become isolated will be small. Leveraging these characteristics, target data is regarded as an anomaly when the average number of splits required to isolate from certain data is smaller than other data.

3.4.4 Random Forest

RF is a classification method based on ensemble learning using decision trees. Multiple training sets are generated from the input data by sampling with replacement, and then each training set is classified based on a decision tree. Importance of the feature quantity is also calculated based on the information obtained from the training stage of RF. In this research, the importance calculated by RF was used in effective feature selection according to defect status.

3.4.5 Laplacian Score

LS is a method to create a neighborhood graph, taking input data as nodes, and perform feature selection based on the graph Laplacian for each neighboring node in the graph. In this research, values of Laplacian score calculated for each feature quantity were used in effective feature selection according to defect status.

4. Evaluation Test

4.1 Test equipment

Fig. 3 shows an outline diagram of the test equipment used for evaluation of defect detection accuracy and **Table 1** shows shapes and sizes of artificial defects created on the test bearings. Rolling bearings (angular contact ball bearing, model No. 7216) were used for the test, with cylindrical holes of different sizes (2b to 8b in **Table 1**) and a rectangular groove sufficiently larger than the cylindrical holes (RG in **Table 1**) created on the outer ring raceway surface. The size of the cylindrical holes was set as 2 to 8 times the minor axis radius (b) of the elliptic elastic contact area of ball and outer ring raceway surface. For example, 2b means the diameter of the cylindrical hole is twice the size of b. These artificial defects assume flaking produced as a result of rolling fatigue, with the cylindrical holes simulating the initial stage of flaking and the rectangular groove simulating the progressed state of flaking.

The test was conducted with a normal bearing without artificial defects (ND in **Table 1**) and bearings with artificial defects of different sizes discussed above, attached to the test equipment. The main shaft of the test equipment was operated at the rotational speeds of 1,000, 1,500 and 2,000 min^{-1} and vibration acceleration in the radial, horizontal, and axial directions was measured and used for evaluation. Evaluation was made individually for each rotational speed of the main shaft. One measurement data of vibration acceleration of a specific direction $\boldsymbol{X}^{(D)} = [x_1\ x_2 \cdots x_i \cdots x_N], D \in \{\text{Axial, Radial, Horizontal}\}$ involved 33 times of measurement for each artificial defect size with sampling frequency of 50 kHz and sampling time of 20 sec. Where, index i indicates an order of time sequence and x_i indicates an instantaneous value of vibration acceleration amplitude of index i. In addition, considering the impact of bearing recombination on vibration, bearings were recombined after 3 measurements.

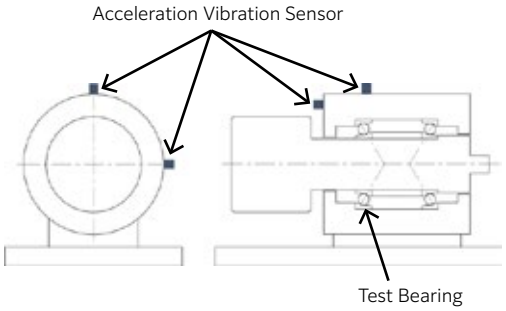


Fig. 3 Test equipment

Table 1 Shape and size of artificial defect

Symbol	Shape of Defect	Size mm
ND	None	-
2b	Hole	ϕ0.32
4b	Hole	ϕ0.64
6b	Hole	ϕ1.02
8b	Hole	ϕ1.36
RG	Rectanguler Groove	Width 2, Height 10, Depth 1

4.2 Calculation of input data v_j

The measurement data of vibration acceleration $X^{(D)}$ was divided into segments at an interval of 5 rotations of the main shaft, and one segment data $y_j^{(D)} = [x_{(j-1)n+1}\ x_{(j-1)n+2} \cdots x_{(j-1)n+k} \cdots x_{(j-1)n+n}]$ was taken for each segment. Where, n is the number of data points included in the segment data, $j = 1,2,3,\cdots, \lfloor N/n \rfloor$ is the segment number and index k indicates the order of time series. Therefore, the total number of segment data $y_j^{(D)}$ which are obtained from a measurement data $X^{(D)}$ is as shown in **Table 2**.

Table 2 Number of segment data y_j in each measurement data X

Rotation Speed	Number of Segment Data
1,000 min^{-1}	66
1,500 min^{-1}	100
2,000 min^{-1}	133

Table 3 Kind of band-pass filter

Filter	Frequency Range(Hz)
Raw	None
Low1	20-200
Low2	20-1,000
Mid1	200-2,000
Mid2	1,000-5,000
High1	2,000-20,000
High2	5,000-20,000

The amplitude data after BPF processing on each segment data $y_j^{(D)}$ in each frequency range shown in **Table 3** was defined as time domain data $y_j^{(D,TIME)}$. In addition, the amplitude data in frequency domain obtained by envelope processing[26] and FFT processing on time domain data $y_j^{(D,TIME)}$ was defined as frequency domain data $y_j^{(D,SPEC)}$ and the amplitude data obtained by another FFT processing on frequency domain data $y_j^{(D,SPEC)}$ was defined as quefrency domain data $y_j^{(D,CEPS)}$. For feature quantities, for $y_j^{(D,R)}$, $R \in$ {TIME, SPEC, CEPS}, statistics such as modulation value (MOF) was used in addition to effective value (OA), maximum value (MAX), crest factor (CF), Kurtosis (KS) and skewness (SKN), which are frequently used, in general, in the diagnosis of rolling bearings. The following shows the calculation method of effective value, maximum value, crest factor, Kurtosis and skewness:

$$OA_j^{(D,R)} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_{j,k}^{(D,R)})^2} \tag{5}$$

$$MAX_j^{(D,R)} = \max_{1\le k\le n} y_{j,k}^{(D,R)} \tag{6}$$

$$CF_j^{(D,R)} = MAX_j^{(D,R)} / OA_j^{(D,R)} \tag{7}$$

$$KS_j^{(D,R)} = \frac{1}{n}\sum_{k=1}^{n}\frac{(y_{j,k}^{(D,R)} - \bar{y}_j^{(D,R)})^4}{(\sigma_j^{(D,R)})^4} \tag{8}$$

$$SKN_j^{(D,R)} = \frac{1}{n}\sum_{k=1}^{n}\frac{(y_{j,k}^{(D,R)} - \bar{y}_j^{(D,R)})^3}{(\sigma_j^{(D,R)})^3} \tag{9}$$

Where, D is the direction of measurement, R is domain, j is the segment number, $y_{j,k}^{(D,R)}$ is the element of index k, $\bar{y}_j^{(D,R)}$ is mean of $y_j^{(D,R)}$, $\sigma_j^{(D,R)}$ is standard deviation of $y_j^{(D,R)}$. Modulation value was defined as the effective value of $y_j^{(D,R)}$ after envelope processing. Also, input data v_j was set as a vectorized feature quantities from different measurement directions and domains in segment data y_j. **Table 4** summarizes the number of parameters included in domain, BPF, statistics and direction of sensor measurement to be considered for calculation of feature quantities. Each input data v_j is composed of total of 378 feature quantities considering domain (3), BPF (7), statistics (6), direction of sensor measurement (3).

Table 4 Number of parameters included in each input data v_j

	Regions	BPFs	Statistics	Sensors	Features
Number of Parameters	3	7	6	3	378

4.3 Selection of feature quantities by variation coefficient

Among various feature quantities included in input data v_j, some quantities exhibit larger variation due to the recombination of bearings rather than variation due to the different sizes of artificial defects. Therefore, variation coefficient (standard deviation divided by mean) was calculated for each feature quantity calculated from segment data y_j of normal bearings, which was used for training, and quantities which variation coefficient exceeded 0.3 were excluded from the target of training and testing, as preprocessing. Approximately 20-40 % of feature quantities, among all the feature quantities calculated in Section 4.2 are excluded by this processing, by the selection of measurement data X used for training. These feature quantities, selected by variation coefficient, were used for training and testing, calculating Z Score for each feature quantity based on the input data v_j of artificial defect size ND, which was used for training described later.

4.4 Selection of training data and test data

Considering the impact of recombination of bearings on defect detection accuracy, a combination of measurement data $\boldsymbol{X}$ was randomly changed for learning and evaluation of the proposed method. **Fig. 4** shows the details of the combination of measurement data $\boldsymbol{X}$, when feature selection is conducted on artificial defect sizes ND and 2b, which is indicated by the red dotted lines in **Fig. 1**. During training, from 33 measurement data $\boldsymbol{X}$ for each artificial defect size, 24 measurement data $\boldsymbol{X}$ were selected for artificial defect sizes ND and 2b, respectively, then the importance of feature quantity was calculated and features were selected by the classification method based on each of the input data $\boldsymbol{v}_j$. The input data after feature selection obtained by this process was set as $\boldsymbol{v'}_j^{2b}$. The superscript 2b indicates that the input data after feature selection $\boldsymbol{v'}_j$ uses a set of feature quantities configured for the artificial defect size 2b. Then, the input data $\boldsymbol{v'}_j^{2b}$ of artificial defect size ND was used for the training outlier detection method, and hyperparameters were selected, so that the specificity is the highest by 10-fold cross validation, and anomaly score a_j^{2b} was calculated.

During the test, anomaly score a_j^{2b} was calculated using the outlier detection method against the input data $\boldsymbol{v'}_j^{2b}$ obtained from 9 measurement data $\boldsymbol{X}$ for each artificial defect size.

Similarly, feature selection was conducted for the combination of artificial defect sizes of ND and 4b, as well as ND and 6b and anomaly scores a_j^{4b} and a_j^{6b} were calculated from the input data after feature selection $\boldsymbol{v'}_j^{4b}$ and $\boldsymbol{v'}_j^{6b}$. In addition, anomaly score $a_j^{(S)}$ against input data after feature selection $\boldsymbol{v'}_j^{(S)}$, $S \in \{2b, 4b, 6b\}$, which can be obtained by this processing, was set as one anomaly score vector $\boldsymbol{a}_j = (a_j^{2b}, a_j^{4b}, a_j^{6b})$ for each corresponding segment. The anomaly score vector $\boldsymbol{a}_j$ against input data $\boldsymbol{v'}_j^{(S)}$ for training/testing was used for training/testing of the second-stage outlier detection method. In addition, we made sure that the measurement data $\boldsymbol{X}$ to be used for training were not duplicated with those for testing.

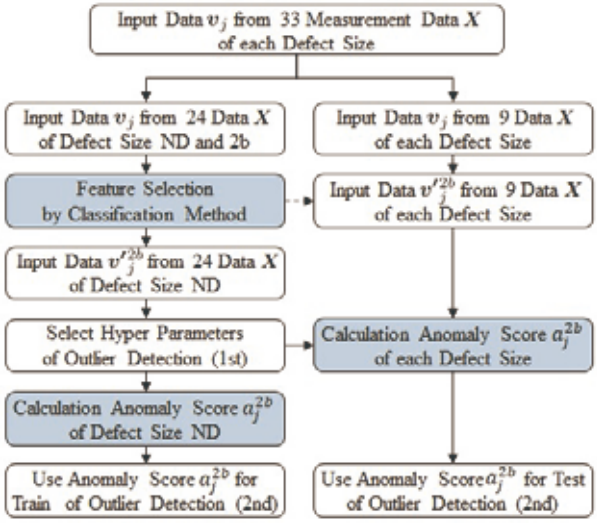


Fig. 4 Selection of measurement data $\boldsymbol{X}$ for train and test

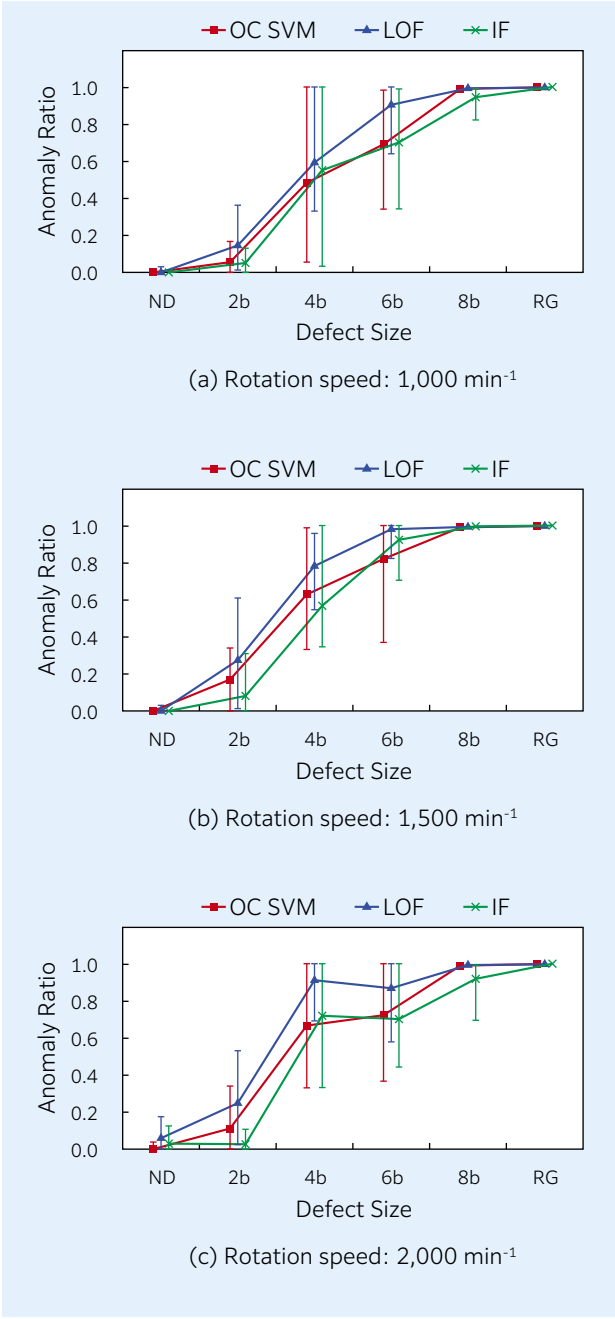
4.5 Testing defect detection accuracy

Defect detection accuracy was tested by calculating the anomaly ratio of each defect size using equation (1) from the anomaly score obtained by the outlier detection method against the measurement data $\boldsymbol{X}$ of the test target. When rotational speed of 1,500 min^{-1} is set as a target, there are 9 measurement data $\boldsymbol{X}$ for test targets of each defect size and 100 segment data $\boldsymbol{y}_j$ obtained from each measurement data, for a total of 900 segment data $\boldsymbol{y}_j$ for the test target of each defect size. The anomaly ratio for each artificial defect size was calculated by using the anomaly score calculated by the outlier detection method against input data $\boldsymbol{v}_j$, $\boldsymbol{v'}_j^{(S)}$ or anomaly score vector $\boldsymbol{a}_j$ from each segment data $\boldsymbol{y}_j$. Measurement data $\boldsymbol{X}$ used for training was randomly changed 50 times and defect detection accuracy was tested by mean and variation of anomaly ratios on differences among the training data.

4.6 Preliminary test

4.6.1 Comparison of outlier detection methods

Fig. 5 shows the relation between artificial defect size and anomaly ratio when feature selection is not made, at rotational speeds of 1,000, 1,500 and 2,000 min^{-1}, for each outlier detection method. In addition, **Fig. 6** shows the result of calculation of AUC Score from different outlier detection methods for artificial defect sizes of ND and 4b, and **Table 5** shows mean, maximum value, minimum value, and standard deviation of hyperparameters obtained by cross validation of different training data. In the table, ν indicates the lower limit of support vector ratio in OC SVM, γ indicates hyperparameter of Gaussian kernel, n-neighbors indicates neighborhood number in calculation of marginal density of LOF, and n-estimators indicates number of estimators in ensemble learning of IF.



(a) Rotation speed: 1,000 min^{-1}

(b) Rotation speed: 1,500 min^{-1}

(c) Rotation speed: 2,000 min^{-1}

Fig. 5 Comparison of outlier detection methods

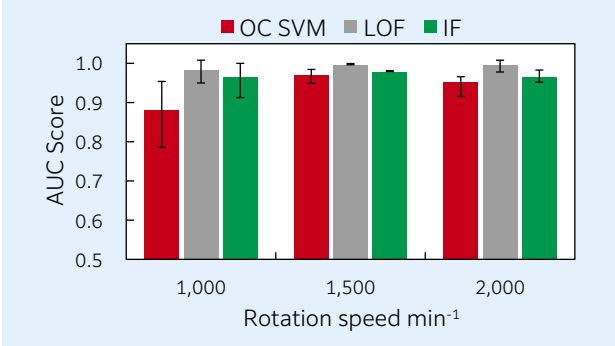


Fig. 6 AUC Score for each method

Table 5 Variation of hyper parameters for difference of train data

Method	Hyper Parameter	Rotation Speed	Average	Max	Min	Standard Deviation
OC SVM	ν	1,000 min^{-1}	2.0×10^{-3}	5.6×10^{-3}	1.8×10^{-4}	2.0×10^{-3}
		1,500 min^{-1}	1.1×10^{-3}	2.0×10^{-3}	4.5×10^{-4}	6.7×10^{-4}
		2,000 min^{-1}	1.6×10^{-3}	5.6×10^{-3}	1.8×10^{-4}	1.5×10^{-3}
	γ	1,000 min^{-1}	2.8×10^{-4}	1.0×10^{-3}	1.0×10^{-4}	2.8×10^{-4}
		1,500 min^{-1}	2.7×10^{-4}	1.0×10^{-3}	7.9×10^{-5}	2.8×10^{-4}
		2,000 min^{-1}	1.6×10^{-4}	3.1×10^{-4}	7.9×10^{-5}	1.1×10^{-4}
LOF	n-neigbors	1,000 min^{-1}	13.3	24	2	8.0
		1,500 min^{-1}	9.8	25	2	8.6
		2,000 min^{-1}	8.4	29	2	8.3
IF	n-estimators	1,000 min^{-1}	33.0	60	17	15.1
		1,500 min^{-1}	32.7	59	23	11.1
		2,000 min^{-1}	71.7	96	35	20.0

Fig. 5 reveals that the average anomaly ratio of artificial defect sizes 2b - 6b is the highest in LOF. **Fig. 6** shows that LOF gives the highest defect detection accuracy for AUC Score also. However, in any method, the average anomaly ratio of artificial defect size 2b is 0.3 or lower, which is lower than the artificial defect size 4b or larger. From **Table 5**, it was estimated that the variation of anomaly ratio was produced along with variation of hyperparameter values depending on different training data. Therefore, feature selection with RF was performed to improve defect detection accuracy.

4.6.2 Calculation of feature importance by Random Forest

Classification accuracy of artificial defect size 2b, which was difficult to detect by the outlier detection method of the previous section, was tested using RF. Initially, in order to confirm the influence of hyperparameter of RF, the number of decision trees of RF at the rotational speed of 1,500 min^{-1} was changed to 100, 1,000 and 10,000, and the maximum depth of decision trees to 10, 100 and 1,000. However, there was no difference in significance level in the influence of artificial defect sizes of ND and 2b on classification accuracy. Therefore, we set the number of decision trees to 1,000 and the maximum depth of decision trees to 100 and conducted evaluation of classification accuracy and feature selection for each rotational speed. **Fig. 7** shows classification accuracy of artificial defect sizes ND and 2b, at rotational speeds of 1,000, 1,500 and 2,000 min^{-1}. **Table 6** shows an example of extracting 10 feature quantities, which are selected based on difference in training data, after feature selection by RF, against artificial defect sizes of ND and 4b;ND and 6b; and ND and 2b at rotational speed of 1,500 min^{-1}. Feature quantities of high importance common to classification of artificial defect sizes ND and 2b and classification of artificial defect sizes ND and 4b are indicated in bold font. Feature quantities of high importance common to classification of artificial defect sizes ND and 4b and classification of artificial defect sizes ND and 6b are indicated in bold font with an underline. Feature quantities are expressed by statistics - domain - BPF (direction of measurement) in the table.

Fig. 7 reveals that artificial defect size 2b can be

classified with very high accuracy by RF regardless of rotational speed. In addition, **Table 6** reveals that 2 feature quantities are common in classification of artificial defect size ND and 2b and classification of ND and 4b, and only 1 feature quantity is common to classification of artificial defect sizes ND and 4b and classification of ND and 6b, while no feature quantity is common to classification of artificial defect sizes ND and 2b and classifications of ND and 6b. Therefore, all the defect sizes cannot be tested with the same feature quantity.

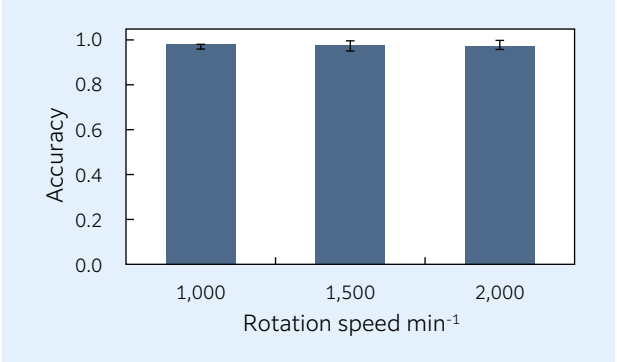


Fig. 7 Classification accuracy by Random Forest

Table 6 Features with high importance to defect size (Rotation speed 1,500 min^{-1})

	Comparison ND and 2b	Comparison ND and 4b	Comparison ND and 6b
1	**OA-TIME-LOW1(Radial)**	**OA-TIME-LOW1(Radial)**	OA-TIME-LOW2(Radial)
2	OA-TIME-LOW1(Horizontal)	OA-CEPS-LOW1(Radial)	MOF-TIME-LOW2(Radial)
3	MAX-TIME-LOW1(Horizontal)	**OA-CEPS-MID1(Radial)**	OA-SPEC-LOW2(Radial)
4	OA-TIME-LOW1(Horizontal)	MAX-CEPS-MID2(Radial)	MOF-SPEC-LOW2(Radial)
5	OA-CEPS-LOW1(Horizontal)	OA-TIME-LOW1(Axial)	OA-CEPS-LOW2(Radial)
6	CF-CEPS-LOW1(Horizontal)	MAX-TIME-LOW1(Axial)	MOF-TIME-MID1(Radial)
7	MOF-TIME-LOW1(Axial)	**MAX-SPEC-LOW1(Axial)**	OA-SPEC-MID1(Radial)
8	OA-SPEC-LOW1(Axial)	SKN-SPEC-LOW1(Axial)	MOF-SPEC-MID1(Radial)
9	**MAX-SPEC-LOW1(Axial)**	OA-CEPS-LOW1(Axial)	**OA-CEPS-MID1(Radial)**
10	MOF-SPEC-LOW1(Axial)	KS-CEPS-LOW1(Axial)	MAX-CEPS-MID1(Radial)

4.6.3 Defect detection accuracy by input data $v'_j{}^{(S)}$ after feature selection

Feature vector composed of the selected 10 feature quantities of high importance with respect to each artificial defect size was set as input data $v'_j{}^{(S)}$, and defect detection accuracy by LOF was tested. **Fig. 8** shows the relation between artificial defect size and anomaly ratio against input data $v'_j{}^{(S)}$ after feature selection. Anomaly ratios calculated from input data $v'_j{}^{2b}$, $v'_j{}^{4b}$ and $v'_j{}^{6b}$ which selected feature quantities of high importance to the classification of artificial defect sizes ND and 2b, ND and 4b, and ND and 6b are shown in the figure as Feature(2b), Feature(4b) and Feature(6b), respectively. In the same figure, anomaly ratios directly calculated from input data v'_j before feature selection using LOF for each defect size are shown as the original method (Original). The results from input data v_j before feature selection are the same as **Fig. 5**. In addition, **Table 7** shows the mean, maximum values, minimum values, and standard deviation of neighborhood numbers in the marginal density calculation of LOF, obtained by cross validation for input data $v'_j{}^{(S)}$ after feature selection.

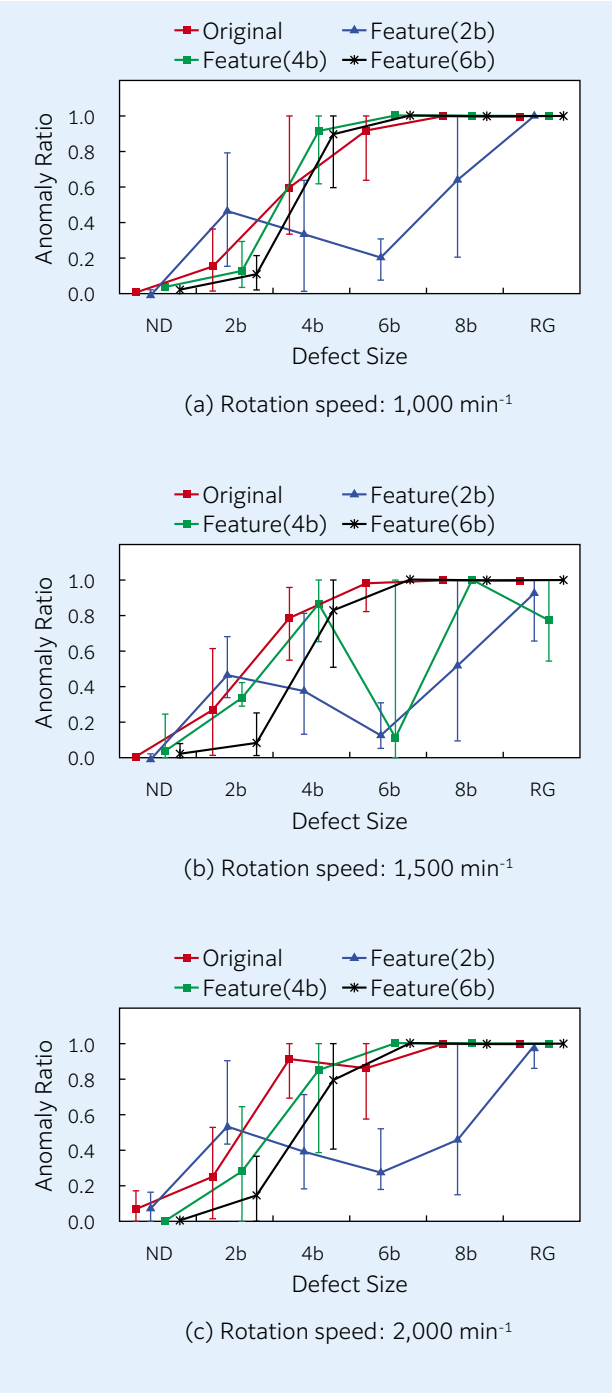


Fig. 8 Anomaly ratio for each input data $v'_j{}^{(S)}$

Fig. 8 reveals that, when feature quantities of high importance are selected for artificial defect size 2b, anomaly ratio of artificial defect size 2b increases compared to the case when feature quantities of high importance for artificial defect sizes 4b and 6b are selected.

On the other hand, when the above feature quantities are selected, anomaly ratios for artificial defect sizes 4b and 6b are lower than anomaly ratios for artificial defect size 2b, and the detection accuracy decreases for those other than the targeted artificial defect size. In addition, **Table 7** shows that variation of the neighborhood number does not

change depending on the different training data. Therefore, defect detection accuracy for each input data $\boldsymbol{v'}_j^{(S)}$ after feature selection must be evaluated comprehensively.

Table 7 Variation of the neighborhood numbers of Local Outlier Factor for the difference of train data

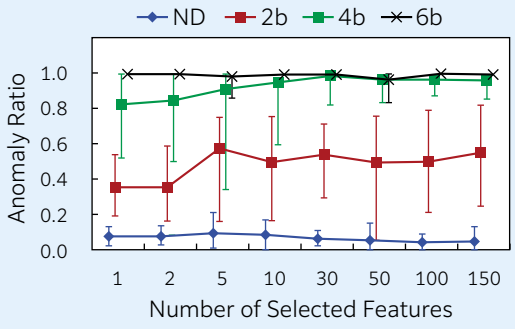
Feature Vector	Rotation Speed	Average	Max	Min	Standard Deviation
Original	1,000 min^{-1}	13.3	24	2	8.0
	1,500 min^{-1}	9.8	25	2	8.6
	2,000 min^{-1}	8.4	29	2	8.3
Feature(2b)	1,000 min^{-1}	16.0	28	3	7.5
	1,500 min^{-1}	14.0	28	5	6.4
	2,000 min^{-1}	24.4	29	16	4.0
Feature(4b)	1,000 min^{-1}	23.1	28	14	4.9
	1,500 min^{-1}	14.4	28	2	8.9
	2,000 min^{-1}	18.1	27	2	7.2
Feature(6b)	1,000 min^{-1}	15.2	28	2	7.8
	1,500 min^{-1}	17.2	27	5	7.0
	2,000 min^{-1}	24.4	29	15	4.6

4.6.4 Outlier detection accuracy by 2-step LOF and impact of number of feature quantities used for feature selection

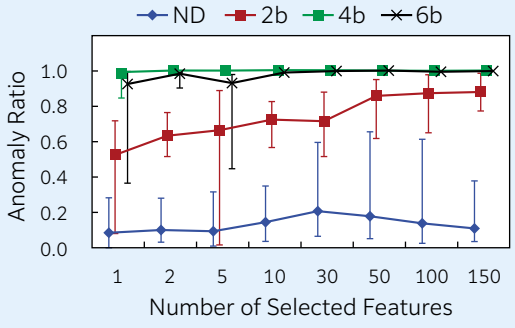
In order to resolve the problem of reduction of anomaly ratio other than artificial defect size targeted for feature selection, the proposed method uses anomaly score for input data $\boldsymbol{v'}_j^{(S)}$ after feature selection as the input data on the second stage of LOF as one vector (anomaly score vector $\boldsymbol{a}_j$) for each segment corresponding to the original measurement data. Then the proposed method compares the defect detection accuracy again. Therefore, defect detection accuracy of the proposed method varies depending on the number of feature quantities in feature selection of input data $\boldsymbol{v'}_j^{(S)}$. **Fig. 9** shows the relation between the number of feature quantities in feature selection of input data $\boldsymbol{v'}_j^{(S)}$ and the mean and variation of anomaly ratios of artificial defect sizes ND, 2b, 4b and 6b in the proposed method at the rotational speeds of 1,000, 1,500 and 2,000 min^{-1}. In addition, **Table 8** shows the relation between the number of feature quantities in feature selection and the mean, maximum values, minimum values, and standard deviation of the neighborhood number.

Fig. 9 shows that the number of feature quantities to improve defect detection accuracy varies depending on the rotational speed. In any rotational speed, the anomaly ratio of artificial defect sizes 2b and 4b decreases when the number of feature quantities is 5 or less. In addition, the anomaly ratio of artificial defect size 2b increases at the rotational speed of 1,500 min^{-1} when the number of feature quantities is 30 or more, however, the anomaly ratio of artificial defect sizes 2b and 6b decreases at the rotational speed of 2,000 min^{-1} and the average anomaly ratio of artificial defect size ND increases up to 0.2. **Table 8** shows no difference in the neighborhood number for different numbers of feature quantities.

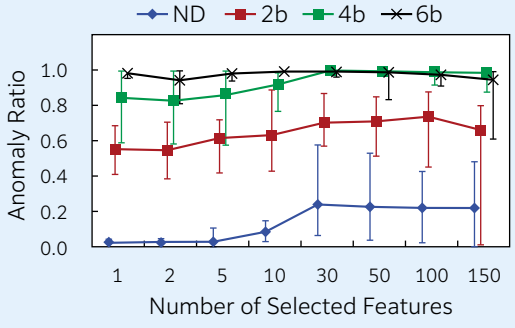
In the practice of defect detection methods, misjudgment of anomaly for normal bearings should be avoided. Also, the larger the difference of anomaly ratios between normal bearings and bearings with small defects, the easier the defect detection will be. Therefore, selection criterion for number of feature quantities adequate for defect detection was determined as the anomaly ratio of artificial defect size ND to be 0.2 or lower and the margin between artificial defect sizes ND and 2b to be large. Furthermore, when the margin between artificial defect sizes of ND and 2b is about the same, features with less numbers of feature quantities should be selected. Based on the above criterion, the number of feature quantities adequate for defect detection was determined to be 30 for rotational speed of 1,000 min^{-1}, 150 for rotational speed of 1,500 min^{-1} and 10 for rotational speed of 2,000 min^{-1}.



(a) Rotation speed: 1,000 min^{-1}



(b) Rotation speed: 1,500 min^{-1}



(c) Rotation speed: 2,000 min^{-1}

Fig. 9 Influence of number of features on feature selection

Table 8 Comparison of the neighborhood numbers of Local Outlier Factor for the number of features

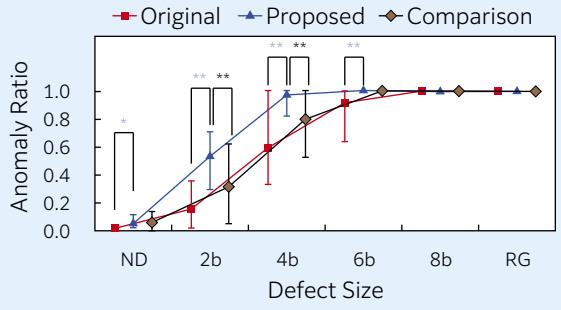
Number of Features	Rotation Speed	Average	Max	Min	Standard Deviation
1	1,000 min^{-1}	11.2	24	3	7.4
	1,500 min^{-1}	14.9	24	6	7.0
	2,000 min^{-1}	16.9	29	8	7.8
2	1,000 min^{-1}	21.5	29	4	8.3
	1,500 min^{-1}	21.3	27	11	5.8
	2,000 min^{-1}	16.2	24	8	7.0
5	1,000 min^{-1}	16.0	25	5	6.5
	1,500 min^{-1}	17.8	29	2	7.9
	2,000 min^{-1}	17.7	29	9	6.7
10	1,000 min^{-1}	16.3	29	6	8.5
	1,500 min^{-1}	18.0	25	3	6.6
	2,000 min^{-1}	22.0	29	16	4.2
30	1,000 min^{-1}	16.9	28	3	10.5
	1,500 min^{-1}	18.1	28	6	6.7
	2,000 min^{-1}	21.8	29	13	5.9
50	1,000 min^{-1}	20.7	29	8	7.4
	1,500 min^{-1}	21.5	29	2	9.4
	2,000 min^{-1}	16.6	28	2	8.5
100	1,000 min^{-1}	15.6	23	8	5.5
	1,500 min^{-1}	16.1	29	2	9.4
	2,000 min^{-1}	21.3	29	2	8.1
150	1,000 min^{-1}	16.6	29	5	7.2
	1,500 min^{-1}	15.6	26	9	5.1
	2,000 min^{-1}	17.7	29	8	7.1

4.7 Defect detection accuracy of the proposed method

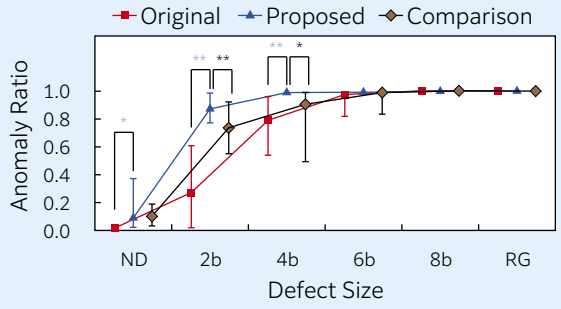
Fig. 10 shows the relation between artificial defect size and anomaly ratio by the proposed method (Proposed) which uses the appropriate number of feature quantities for defect detection selected by the previous section at rotational speeds of 1,000, 1,500 and 2,000 min^{-1}. As a comparison, anomaly ratios directly calculated from input data $\boldsymbol{v}_j$ before feature selection using LOF for each artificial defect size are shown as the original method (Original). Results of the original method are the same as **Fig. 5**. Also, the anomaly ratio for each artificial defect size using Laplacian Score[16] as the feature selection method is depicted in the figure as the comparison method (Comparison). **Fig. 11** shows the comparison of AUC Score among the proposed method, original method and comparison method against artificial defect sizes ND and 2b. In the anomaly ratio of **Fig. 10** and AUC Score of **Fig. 11**, * is marked on the difference of significance level of 5 % between the original method and the proposed method, and the comparison method and the proposed method, and ** is marked on the difference of significance level of 1 %. **Table 9** shows the mean, maximum value, minimum value, and standard deviation of the neighborhood numbers of the proposed method and original method.

Fig. 10 indicates that the anomaly ratios of artificial defect size 6b or more by the proposed method are almost 1 regardless of rotational speed and even those for artificial defect size 2b significantly increased compared to the original method and comparison method. Also, for artificial defect size 4b, the anomaly ratio by the proposed method significantly increased at rotational speeds of 1,000 and 1,500 min^{-1} over the original method, and regardless of rotational speed over the comparison method. From **Fig. 11**, AUC Score can be improved by using the proposed method at any rotational speed over the original and comparison methods. At rotational speeds of 1,000 min^{-1} and 1,500 min^{-1}, although the anomaly ratio of artificial defect size ND shows significant difference between the original method and the proposed method, AUC Score also significantly improved with the proposed method. Therefore, use of the proposed method improved the defect detection accuracy.

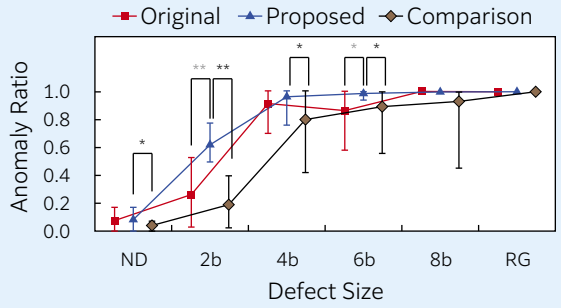
Table 9 reveals no difference in variation of neighborhood number of LOF between the original and the proposed methods.



(a) Rotation speed: 1,000 min^{-1} (30 Features)



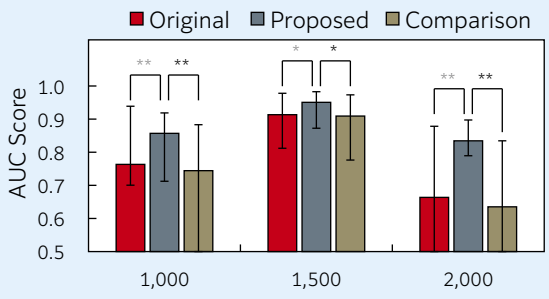
(b) Rotation speed: 1,500 min^{-1} (150 Features)



(c) Rotation speed: 2,000 min^{-1} (10 Features)

* : Significance level 5 %
** : Significance level 1 %

Fig. 10 Comparison of anomaly ratios between original, comparison and proposed method



* : Significance level 5 %
** : Significance level 1 %

Fig. 11 Comparison of AUC Score between original, comparison and proposed method

Table 9 Comparison of the neighborhood numbers of Local Outlier Factor between original and proposed method

Method	Rotation Speed	Average	Max	Min	Standard Deviation
Original	1,000 min^{-1}	13.3	24	2	8.0
	1,500 min^{-1}	9.8	25	2	8.6
	2,000 min^{-1}	8.4	29	2	8.3
Outlier Detection (2nd) (Proposed)	1,000 min^{-1}	16.3	29	6	8.5
	1,500 min^{-1}	18.0	25	3	6.6
	2,000 min^{-1}	22.0	29	16	4.2

5. Future challenges

The test equipment used for this evaluation showed large variation due to recombination of bearings. It is considered that recombination has some impact on the test environment. For validation of the proposed method, evaluation with completely different test equipment is necessary. Currently, data collection using new test equipment is underway, with a plan to validate the proposed method moving forward.

In the proposed method, feature selection with supervised learning is conducted for calculating anomaly score in the first stage of LOF and vibration data of bearings with defects are used as supervised data. However, in practice, it is almost impossible to obtain vibration data of bearings with defects. As a solution to this problem, for example, a method to improve defect detection accuracy by extracting features which do not depend on rotational speed or equipment can be considered, using vibration analysis methods such as Persistent Homology[27] and Dynamic Mode Decomposition[28]. Or using data obtained from a theoretical analysis model instead of measurement data from the actual equipment as the supervised data can be considered, by using the theoretical analysis model[29] such as Dynamic Analysis in feature

selection by Random Forest. Or, using Transfer Learning, data from test equipment or a theoretical analysis model may be used for diagnosis of other equipment as supervised data. We will study these possibilities for improvement.

6. Summary

Various anomaly detection methods were compared for defect detection accuracy against vibration acceleration data of bearings with artificial defects. Insights obtained from this study are as follows:

- When 3 types of outlier detection methods, One Class Support Vector Machine, Local Outlier Factor and Isolation Forest, were compared for defect detection accuracy, Local Outlier Factor provided the highest defect detection accuracy. However, in all methods, defect detection accuracy of artificial defect size 2b was lower than artificial defect size 4b or larger.
- Evaluation of classification accuracy of artificial defect size by Random Forest revealed that feature quantities of high importance for classification of bearings with and without defects are different depending on the size of artificial defects. In addition, feature selection by Random Forest for micro artificial defect size improved the defect detection accuracy of the targets.
- The proposed defect detection method successfully provided a significant improvement of defect detection accuracy for small artificial defect size without reducing detection accuracy of larger defect size, compared to the cases of no feature selection by Random Forest.

Acknowledgment

This paper was produced by editing the original paper (Kitai, Akamatsu and Fukui, Defect Detection Method for Rolling Bearing Including Micro Defect by Feature Selection and Two Step Outlier Detection Method, Transactions on Mathematical Modeling and its Applications, The Information Processing Society of Japan Vol. 12, (2019) pp 32-42). We appreciate the generosity of the Information Processing Society of Japan.

References

1) Akio Igarashi, Banda Noda, Eichi Matsushima, A Study on Fault Prediction of Rolling Bearings (1st Report), Journal of Japan Society of Lubrication Engineers, Vol. 24, No. 2, (1979) 122–129.
2) Akio Igarashi, Hiroyoshi Hamada, Research on Vibration and Sound of Defective Rolling Bearings (1st Report), Transactions (C) of The Japan Society of Mechanical Engineers, Vol. 47, No. 422, (1981) 1327-1336.
3) Hiroki Mano, Atsushi Korenaga, Diagnosis for rolling bearing using acoustic emission and vibration technique, Proceedings of JSPE Spring Conference of The Japan Society for Precision Engineering, (2014) 683-684.
4) Schölkopf, B., Platt, J.C., Taylor, J.S., Smola, A.J., and Williamson, J., Estimating the Support of a High Dimensional Distribution, Neural Computation, Vol. 13, (2001) 1443–1471.
5) Takashi Onoda, Norihiko Ito, Hideaki Koreeda, Data Indicating Malfunction Sign of Hydroelectric Power Plants, The transactions of the Institute of Electrical Engineers of Japan. D, Vol. 131, No. 4, (2011) 448–457.
6) Tax, D.M.J., One Class Classification, Ph.D thesis, Delft University of Technology (2001).
7) Minoru Kondo, Tatsuro Takashige, Shin'ichi Manabe, Hiroshi Kanno, Abnormality Detection in a Contaminated Diesel Engine by a Vibration Monitoring Method, RTRI Report, Vol. 30, No. 4, (2016) 47–52.
8) Hermansky, H., Ellis, D.P. and Sharma, S., Tandem Connectionist Feature Extraction for Conventional HMM Systems, IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings, Vol. 3, (2000) 1635–1638.
9) Takanori Hasegawa, Jun Ogata, Masahiro Murakawa, Tetsuji Ogawa, Advancement of anomaly detection technology for wind turbine system based on representation learning for normal and faulty classification, 39th Symposium on Wind Energy Application (2017)
10) Masashi Kitai, Hideyuki Tsutsui, Anomaly Diagnosis of Angular Ball Bearings using One Class Support Vector Machine, Proceedings for Tribology Conference 2017 Fall (2017) C41.
11) Koma Kato, Ryoji Tani, Hideyuki Tsutsui, Relationship between defect size and various vibration features in angular contact ball bearings, Proceedings for Tribology Conference 2017 Spring (2017) F33.
12) Breunig, M.M., Kriegel, H.P., Ng, R.T., and Sander, J., LOF, Identifying Density-Based Local Outliers, Management of Data, Vol. 29, (2000) 93–101.
13) Breiman, L., Random Forests, Machine Learning, Vol. 45, (2001) 5–32.
14) Smith, J.S., The Local Mean Decomposition and Its Application to EEG Perception Data, J.R.Soc. Interface, Vol. 2, (2005) 443-445.
15) Aziz, W. and Arif, M., Multiscale Permutation Entropy of Physiological Time Series, 9th International Multitopic Conference, (2005) 1-6.

16) He, X., Cai, D. and Niyogi P., Laplacian Score for Feature Selection, Adv.Neural Inform. Process. Syst., (2005) 16.
17) Cheong, S., Sang, H.O. and Lee, S.Y., Support Vector Machines with Binary Tree Architecture for Multiclass Classification, Neural Inform. Process. -Lett. Rev., Vol. 2, (2004) 47-51.
18) Li, Y., Xu, M., Wei,Y., and Huang, W., A New Rolling Bearing Fault Diagnosis Method Base on Multiscale Permutation Entropy and Improved Support Vector Machine Based Binary Tree, Measurement, Vol. 77, (2016) 80-94.
19) Kilundu, B., Chiementin, X. and Dehombreux, P., Singular Spectrum Analysis for Bearing Defect Detection, Journal of Vibration and Acoustics, Vol. 133, No. 5, (2011) 051007.
20) Bugharbee, H.A. and Trendafilova, I., A New Methodology for Fault Detection in Rolling Element Bearings using Singular Spectrum Analysis, The International Journal of Condition Monitoring, Vol. 7, No. 2, (2018) 26-35.
21) Zhu H., Wang Y., and Wang K., Particle Swarm Optimization (PSO) for the Constrained Portfolio Optimization Problem, Expert System, Vol. 38, (2006) 10161-10169.
22) Hinton G. E. and Osindero S., A Fast Learning Algorithm for Deep Belief Nets, Neurocomputing. Vol. 18, (2006) 1527-1554.
23) Shao, H., Jiang, H., Zhang, X., and Niu, M., Rolling Bearing Fault Diagnosis Using an Optimization Deep Belief Network, Measurement Science and Technology, Vol. 26, (2015) 1-17.
24) Tsuyoshi Ide, Masashi Sugiyama, Anomaly Detection and Change Detection, Kodanshsa, (2015) 11–12.
25) Liu, F.T., Ting, K.M. and Zhou, Z.H., Isolation Forest, Eighth IEEE International Conference on Data Mining, (2008) 413–422.
26) Toshio Toyota, Procedure of anomaly diagnosis of rotating machines, Japan Institute of Plant Maintenance, (1991) 94–96.
27) Umeda Y., Time Series Classification via Topological Data Analysis, Transactions of the Japanese Society for Artificial Intelligence, Vol. 32, No. 3, (2017) 1-12.
28) Kota Dohi, Naoya Takeishi, Takehisa Yairi, Koichi Hori, Anomaly detection of acoustic data using dynamic mode decomposition, 32nd National Conference of the Japanese Society for Artificial Intelligence, 1P2-02, (2018).
29) Gupta, P.K., Advanced Dynamics of Rolling Elements, Springer-Verlag, (1984).

Photo of authors

Masashi KITAI
Product and Business Strategic Planning Dept., New Product and Business Strategic Planning Headquarters

Yoshinobu AKAMATSU
New Product and Business Strategic Planning Headquarters

Ken-ichi FUKUI
SANKEN (The Institute of Scientific and Industrial Research), Osaka University