Smart Structures and Systems, *Vol. 29, No. 6 (2022) 757-766*
https://doi.org/10.12989/sss.2022.29.6.757

Normal data based rotating machine anomaly detection using CNN with self-labeling

Jaewoong Bae[a], Wonho Jung[b] and Yong-Hwa Park*

Department of Mechanical Engineering, Korean Advanced Institute for Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

(*Received October 31, 2021, Revised January 31, 2022, Accepted April 26, 2022*)

Abstract. To train deep learning algorithms, a sufficient number of data are required. However, in most engineering systems, the acquisition of fault data is difficult or sometimes not feasible, while normal data are secured. The dearth of data is one of the major challenges to developing deep learning models, and fault diagnosis in particular cannot be made in the absence of fault data. With this context, this paper proposes an anomaly detection methodology for rotating machines using only normal data with self-labeling. Since only normal data are used for anomaly detection, a self-labeling method is used to generate a new labeled dataset. The overall procedure includes the following three steps: (1) transformation of normal data to self-labeled data based on a pretext task, (2) training the convolutional neural networks (CNN), and (3) anomaly detection using defined anomaly score based on the softmax output of the trained CNN. The softmax value of the abnormal sample shows different behavior from the normal softmax values. To verify the proposed method, four case studies were conducted, on the Case Western Reserve University (CWRU) bearing dataset, IEEE PHM 2012 data challenge dataset, PHMAP 2021 data challenge dataset, and laboratory bearing testbed; and the results were compared to those of existing machine learning and deep learning methods. The results showed that the proposed algorithm could detect faults in the bearing testbed and compressor with over 99.7% accuracy. In particular, it was possible to detect not only bearing faults but also structural faults such as unbalance and belt looseness with very high accuracy. Compared with the existing GAN, the autoencoder-based anomaly detection algorithm, the proposed method showed high anomaly detection performance.

Keywords: anomaly detection; convolutional neural network; deep learning; pretext task; self-labeling

1. Introduction

Rotating machines are widely used in industrial applications such as power plants, manufacturing machines, high-speed trains, and electric vehicles. Unexpected faults in most engineering systems can lead to undesired accidents, increased maintenance costs, reduced machine reliability, and human losses. It is important to detect faults in machines so that they can be operated safely and efficiently. The characteristics of rotating machines are affected by their operating conditions, manufacturer, maintenance history, and capacity. Therefore, an optimized fault detection method is required for accurate detection.

Fault detection methods can be categorized into three types: model-based, data-driven, and signal-based methods (Dai and Gao 2013, Isermann 1984). The first, model-based fault detection method simulates target system based on domain knowledge and estimates the physical parameters related to faults. If the target system is complex, a lot of time and effort is required in the system modeling process. In addition, real-time fault detection is very difficult due to the large computational burden.

The second, data-driven method uses information from a large number of data measured by the target system (Choi *et al*. 2012, Yin *et al*. 2014). Compared with the model-based method, the data-driven method requires less domain knowledge and expertise. Therefore, the data-driven method is effective for fault detection in a complex engineering system. To develop a high-performance fault detection method, a lot of healthy data and faulty data are needed. However, it is difficult to collect a large number of fault data due to the regular maintenance and improvement of the system.

Finally, the signal-based method detects faults by analyzing the signals associated with the fault (da Silva *et al*. 2008, Srividya *et al*. 2009). The type of signal is determined by the operating environment and the type of fault. The type of signals can be generally divided into stationary and non-stationary, as shown in Fig. 1 (Randall 2010). The stationary signals can be further categorized into random, periodic, and quasi-periodic. Periodic and quasi-periodic signals are deterministic and made up of discrete sinusoidal components. The non-stationary signals can be grouped into transient, cyclostationary, and continuously varying. Cyclostationary signals are amplitude-modulated white noise. To capture the characteristics of the signals, appropriate signal processing techniques should be used

*Corresponding author, Professor
E-mail: yhpark@kaist.ac.kr
[a]M.S. Student
E-mail: jwoong.bae@kaist.ac.kr
[b]M.S.
E-mail: wonho1456@kaist.ac.kr

ISSN: 1738-1584 (Print), 1738-1991 (Online)

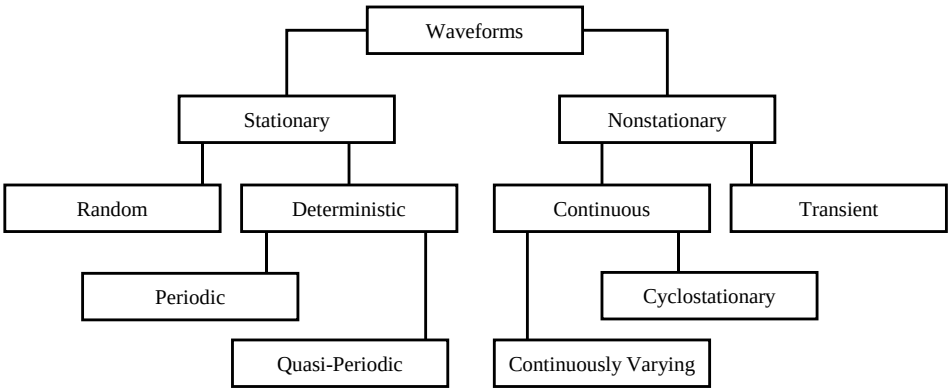


Fig. 1 Signal type

depending on the type of signals. However, it is challenging to select the most relevant signal technique since the dynamics of an engineering system are often complex and multiple signal types exist in data collected from the system.

Many signal processing methods have been proposed in effort to adapt data-driven methods to signal processing. A feature that indicates the health condition of the target system can be designed by analyzing signals measured from the system in time, frequency, and time-frequency domains, based on domain knowledge. Typical signal processing methods, including fast Fourier transform (FFT), short-time Fourier transform (STFT), wavelet transform (WT), and Wigner-Ville distribution (WVD), can be utilized to monitor health conditions of systems. However, those methods require a large effort to select the best signal processing method depending on machine type and environment.

With recent advances in sensor, storage, and computing power, higher performance data-driven fault detection has emerged that is less time-consuming, and requires less expertise. Hoang and Kang (2019) suggested a method that classifies bearing defects using CNN and vibration data under various conditions. Tao *et al*. (2015) used stacked multiple autoencoders to extract fault features from raw vibration signals and perform softmax regression to recognize fault type. The deep learning-based detection showed higher performance than the conventional method. However, the above-mentioned deep learning algorithm was not able to achieve high performance under actual field conditions because of a data imbalance problem.

The data imbalance problem is a major issue in prognostics and health management (PHM) techniques. To solve this problem, data augmentation techniques such as generative adversarial networks (GAN), and autoencoder (AE) have been used. However, in practice, a number of data are not sufficient and many cases are not labeled. And since data augmentation methods eventually require distribution of the faulty data, having insufficient data makes it difficult to generate the distribution of faulty data.

To overcome this limitation, an anomaly detection algorithm has been employed. Anomaly detection is a method that detects abnormal samples that are outliers from the normal sample distribution. This method can be applied in various fields, for example, distinguishing between normal and disease tissues in medical images, detecting spam emails using text data, and detecting abnormal consumption patterns in financial fraud.

The anomaly detection algorithm is divided into two types depending on how the data are labeled: (1) supervised learning, which uses a dataset with labeled data for normal and abnormal classes, and (2) semi-supervised, which uses labeled data for the normal class only. Notably, semi-supervised anomaly detection emerges due to the difficulty in acquiring fault data in the real field. Deep learning-based anomaly detection using only normal samples has also been studied. Semi-supervised anomaly detection can be divided into three types: (1) health-index based anomaly detection, (2) GAN based anomaly detection, and (3) auto encoder (AE) based anomaly detection.

Health-index based anomaly detection is a method of analyzing data and machine characteristics, choosing health indexes for fault detection using clustering (Tian *et al*. 2015) and machine learning algorithms such as support vector machine (Guo *et al*. 2009). These methods require a lot of experience and time for index selection, and if a meaningless index is selected due to insufficient analysis of the fault, the detection accuracy decreases.

GAN, suggested by Goodfellow *et al*. (2014), is a generative adversarial neural network that uses a generator to create learned data samples from latent space z. GAN learns the distribution of the normal data during the training process; the abnormal detection is then made using the difference between distributions of the normal data and the abnormal data. Anomaly detection algorithms using GAN include AnoGAN (Schlegl *et al*. 2017), and GANomaly (Akcay *et al*. 2018). Since the GAN model solves the Nash equilibrium problem between the discriminator and the generator, problems such as mode collapse and model oscillation can occur, which makes training difficult.

AE consists of an encoder that compresses data and a decoder that decodes it to generate inputs from the compressed data. After training, the encoder acts as a feature extractor. Autoencoder-based anomaly detection trains the autoencoder using normal data, and the trained encoder extracts the main features of the normal data. Then, anomaly detection is performed using the extracted features. Since the autoencoder uses both encoder and decoder models at the same time, a number of parameters increases.

Self-supervised learning has recently been used to learn

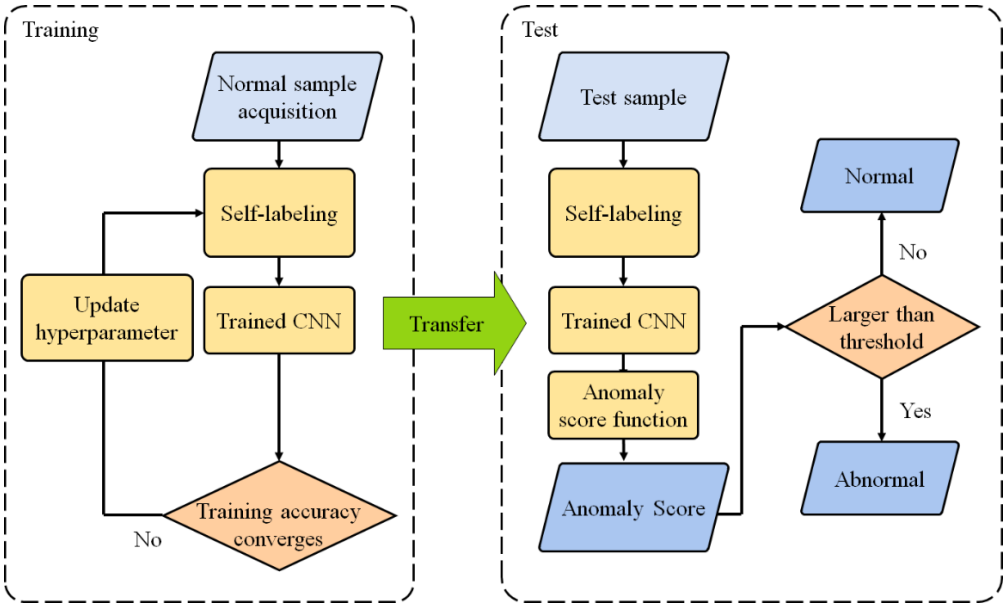


Fig. 2 Framework of the proposed method for fault detection

representations of data, and has exhibited high performance for learning data representations. We propose a novel anomaly detection algorithm by introducing the concept of pretext tasks used in self-supervised learning and CNN. The remainder of this paper is organized as follows. Section 2 provides a brief overview of the CNN network and pretext task used in our method. Section 3 introduces a new anomaly detection algorithm using the CNN network and pretext task concept. In Section 4, a case study is conducted on abnormal detection and early fault detection experiments. Finally, the Conclusions are presented in Section 5.

2. Related work

2.1 Convolutional neural network

A CNN is an artificial neural network that utilizes convolution operations. By using convolution operations, the spatial information of 2D data can be maintained and sent to the next layer (LeCun *et al*. 1998). CNN has shown high performance in computer vision tasks such as image classification (Krizhevsky *et al*. 2012), object detection (Girshick *et al*. 2014), and segmentation (Shelhamer *et al*. 2017).

CNN networks have demonstrated high classification performance for vibration data. To take advantage of the strengths of CNN, a 2D imaging method using raw vibration data is proposed. The 2D imaging method alters 1D time-series data into a 2D image that emphasizes fault characteristics (Amar *et al*. 2015, Hoang and Kang 2019, Oh *et al*. 2018, Wang *et al*. 2019). Jia *et al*. (2018) analyzed a convolution layer trained using bearing fault diagnosis data, and it was confirmed that the learned convolutional layer acted as a bandpass filter to extract the specific frequency bands required for detection. Those results suggest that CNN has high performance in fault diagnosis.

Beyond the recent improvements in diagnosis accuracy, CNN has also been used to predict the remaining useful life (RUL) of a bearing. In the RUL prediction, the remaining lifetime is derived by assuming that the failure occurs at the time the predicted health index (HI) reaches a threshold. HI is defined as the softmax value of CNN (Yoo and Baek 2018).

The reason why CNN networks are excellent in solving such problems is that they automatically learn features required for tasks during the training process. According to Zeiler and Fergus (2014), simple features are extracted in the low-level layer, and more complex patterns are extracted in the high-level layer. In addition, since a CNN network has the shared-weight architecture of a convolutional layer, it can extract translation-invariant features, and so even if the position of the input changes, the same characteristics are extracted (Ren *et al*. 2017).

2.2 Pretext task

In general, a model using supervised learning can guarantee high performance, but it requires a labeled dataset. However, it takes a lot of time and effort to build a labeled dataset. To reduce the time and effort required for data annotation, self-supervised learning has been widely used to learn features from unlabeled data. In self-supervised learning, a neural network learns in two steps: the pretext task and downstream (Jing and Tian 2020). The pretext task is a pre-designed task for networks to solve, and while solving the pretext task network learns to extract features from data. In the computer vision field, the pretext is carried out in various ways: for example, by coloring a gray image (Zhang *et al*. 2016), and finding the original arrangement of shuffled patch images (Noroozi and Favaro 2016). Golan and El-Yaniv (2018) detected anomalies using a geometric transformation and pretext task in an image, it showed higher anomaly detection performance than other methods.

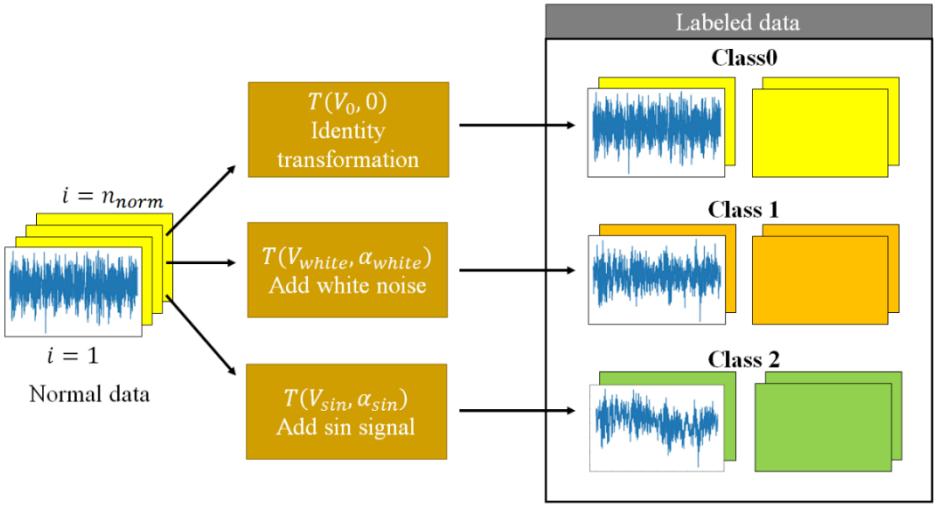


Fig. 3 Generate labeled dataset using self-labeling

3. Proposed method

3.1 Overall procedure

The overall procedure for anomaly detection is shown in Fig. 2. In the training step, the first normal data are collected from a normal state machine for training, then the collected data are segmented with constant length to make a normal dataset. Segmentation length is set to capture one revolution of the rotating machine as in Eq. (1) (Meng *et al.* 2019), where *seg_len* is segmentation length, *Fs* is the sampling frequency, and *RPM* is the rotating speed of the system. To learn the representation of normal data in a classification manner, self-labeling method is introduced. A self-labeling method creates a new labeled dataset using only existing normal data. The CNN network is trained using the generated labeled dataset, and the hyperparameters of self-labeling are updated until training accuracy converges (Section 3.3). In the test step, to quantify the abnormality of the data, a new anomaly score function is defined. A threshold is set through the distribution of the anomaly scores of the normal data, and if the anomaly score of the test sample x is greater than the threshold, it is regarded as abnormal (Section 3.4).

$$sample\ \ length = Fs * 60 / RPM \tag{1}$$

3.2 Self-labelling

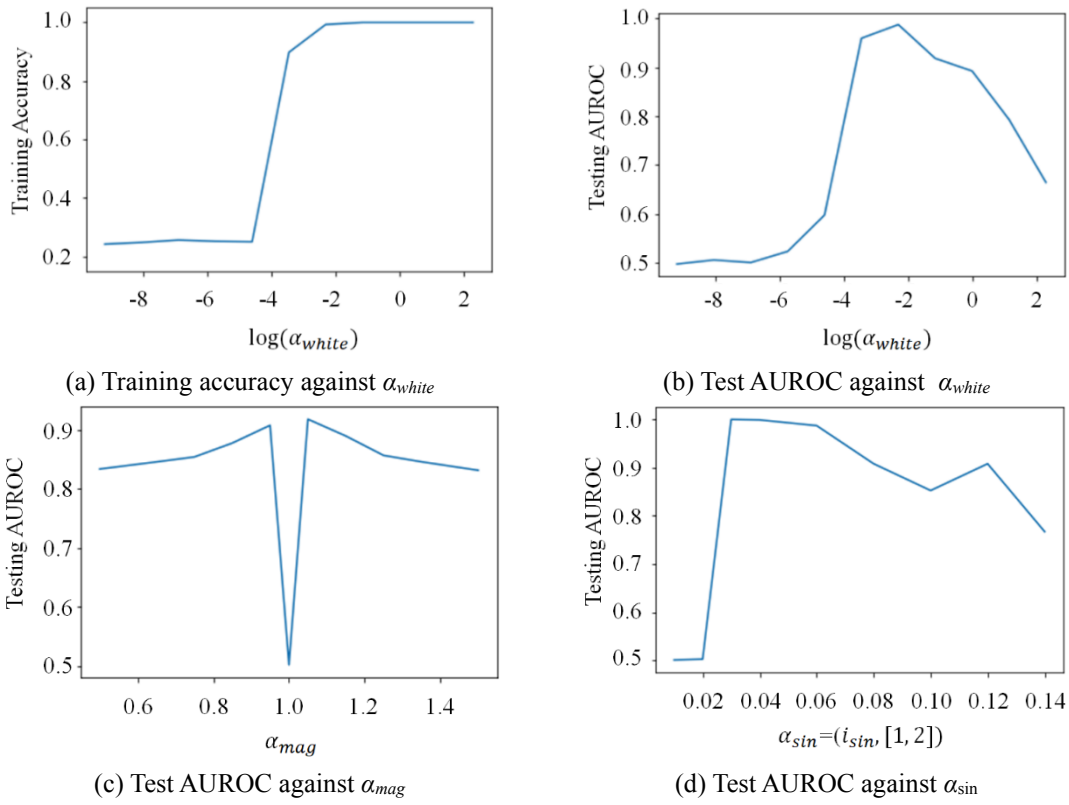
To learn representations of normal data, the pretext task concept is introduced to train the CNN network. Vibration data are time-series data that have a sequence according to time, and have the characteristic of oscillating with a specific frequency. Based on the above information, a pretext task suitable for vibration data is defined. The designed pretext task is to classify the labeling dataset generated by applying normal data to self-labeling. Self-labeling is a technique to create a newly labeled dataset by adding a signal or modulating a raw vibration signal. The process of creating a labeled dataset using the self-labeling method is shown in Fig. 3.

Let S be a normal data sample and $D = \{D_0, D_1, D_2, \cdots D_{n-1}\}$ be transformation, D_0 is the identity transformation. The data sample created by the D_i is S_i, and these data are labeled as class i. The generated dataset by D is denoted as S_D and has $S * n$ data and n classes. In the relation $D_i = T(V_j, \alpha_k)$, the index i is determined by the variation method V_j and the hyperparameter α_k. Even if the same variation method is used, a different index i is assigned if the hyperparameter is different. V_j is a variation method that adds a signal to normal data or modifies normal data. Three methods $V = \{V_{white}, V_{mag}, V_{sin}\}$ including white noise (V_{white}), magnitude variation (V_{mag}), and harmonic frequency addition (V_{sin}) are used.

The three variation methods simulate the background noise or external noise transmitted from other mechanical systems during the vibration data acquisition process. The reason for choosing these variation methods is to create discriminable data while maintaining the characteristics of the normal data when an external signal is added, and to train the CNN network to extract features robust to noise.

3.3 Hyperparameter optimization

The self-labeling method requires hyperparameters, and hyperparameters affect the anomaly detection performance. An experiment was conducted to determine how the hyperparameter affects the training and the classification performance in the test stage. The change in training accuracy was checked by training the CNN on the labeled dataset S_D, created by applying transformation $D = \{D_0, D_1\}$ to the normal data sample S. D_0 is the identity transformation and $D_1 = T(V_{white}, \alpha_{white})$ is the white noise addition method with hyperparameter α_{white} which determines the intensity of the white noise. In the test step, the anomaly score is calculated for the normal sample and the fault sample, then the AUROC value is used to compare the performance according to the change in alpha. The AUROC value is the area under the Receiver Operating Characteristics (ROC). It is a performance metric used to evaluate the classification performance of a model. The



(a) Training accuracy against α_{white} (b) Test AUROC against α_{white}

(c) Test AUROC against α_{mag} (d) Test AUROC against $\alpha_{\sin}$

Fig. 4 Accuracy, AUROC against α

ROC consists of TPR (True Positive Rate) on the x-axis and FPR (False Positive Rate) on the *y*-axis. AUROC has a value between 0 and 1, and the closer to 1, the better the model.

The training accuracy and testing AUROC were plotted against $log\ (\alpha)$, and the results are shown in Fig. 4. First, regarding the training accuracy, as α_{white} increased, the training accuracy increased. Training did not perform well at a small α_{white} value, because the difference between the normal data D_0 and D_i generated by $T(V_{white},\ \alpha_{white})$ cannot be distinguished by the network. The changes in testing AUROC against $log\ (\alpha)$ are shown in Fig. 4(b). At a small α_{white} value, the AUROC value increases as the training accuracy increases, but when the $log(\alpha)$ becomes larger than -2, the AUROC value decreases. The reason AUROC decreases is because when α_{white} becomes larger, the noise signal becomes larger than the original normal signal, and the network learns to distinguish the difference between noises that are not normal signals characteristics.

In the case of magnitude variation, the test AUROC change was confirmed for $D = \{D_0, T(V_{mag}, \alpha_{mag})\}$. α_{mag} is constant to be multiplied by the raw vibration signal. When α_{mag} moved away from 1, AUROC reached a maximum and then gradually decreased. In the case of harmonic frequency addition with $\alpha_{sin} = (i_{sin}, o_{sin})$ which determines the intensity of the sinusoidal signal and the order of harmonic, the AUROC value increased with increasing alpha at small α_{sin}. But after convergence, AUROC showed a tendency to decrease as alpha increased. Through this, the α_{white}, α_{sin} could be updated using Eq. (2). For α_{mag} Eq. (2) was used when α_{mag} >1, and Eq. (3) was used when α_{mag} was smaller than 1.

$$\alpha = \alpha + learning\_rate*(1-accuracy_{train}) \quad (2)$$

$$\alpha = \alpha - learning\_rate*(1-accuracy_{train}) \quad (3)$$

3.4 Anomaly score

The anomaly score function was defined in order to quantify the anomality of the test sample. The anomaly score is a number that quantifies how much a test sample x is different from the trained normal sample. Let $D = \{D_0, D_1, D_2, \cdots D_{n-1}\}$ be the fixed transformation determined by the hyperparameter optimization mentioned in Section 3.3, and S is the training dataset consisting of only normal data. F is the n-class classifier CNN trained using S_D. $F(x)$ is the $1 \times n$ vector of softmax output of F applied on x. $F(x|i)$ is i th $F(x)$ which represents the probability that x is included in the i class. $F(D_i(x)|i)$ indicates the probability that the data $D_i(x)$ generated by applying D_i to x will be included in class i, and the anomaly score function is defined as in Eq. (4). The anomaly score is the sum of the probability that the data $D_i(x)$ is not classified in the i class.

In the next step, a threshold is set for distinguishing normal data from abnormal data. Assuming that the normal sample follows a normal distribution, a threshold is determined to be normal if it is within the 95% confidence interval. This confidence interval can be set differently depending on the task. The threshold is defined as Eq. (5). If the anomaly score was greater than this value, it was

Table 1 Seeded fault type and size

Fault type	Fault size
Ball fault	0.007, 0.014, 0.021 inch
Inner race fault	0.007, 0.014, 0.021 inch
Outer race fault	0.007, 0.014, 0.021 inch

Table 2 Accuracy for every fault type and size

Crack size	Ball fault	Inner race fault	Outer race fault
0.007 inch	100%	100%	100%
0.014 inch	100%	100%	100%
0.021 inch	100%	100%	100%

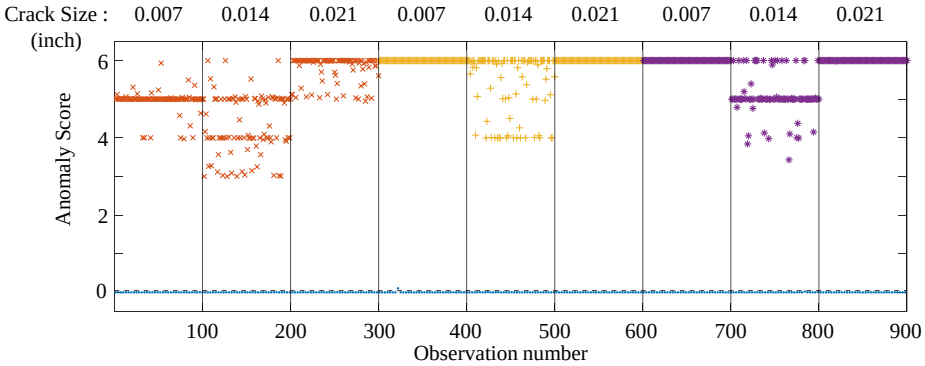


Fig. 5 Anomaly scores for normal and fault samples (normal sample: blue dot, ball fault: red X, yellow cross: Inner race fault, purple asterisk: Outer race fault)

judged to be abnormal. λ is the threshold, $\bar{S}$, σ are the mean and standard deviations of the anomaly score of normal data, and $z_{0.05}$ is the z-value for the 95% confidence interval.

$$A(x) = \sum_{i=0}^{n-1} (1 - F(D_i(x_i) \mid i)) \tag{4}$$

$$\lambda = \bar{S} + z_{0.05} * \sigma \tag{5}$$

4. Case study

To verify the proposed method, four case studies were conducted. The CWRU bearing dataset was described in Section 4.1 and is used to conduct anomaly detection for different severities of bearings and type of faults (CWRU Bearing Data Center 2022). In Section 4.2, verification was carried out using the data used in the PHMAP2021 data challenge (Data challenge on PHMAP 2021). These data were collected from an oil injection screw compressor and it has three different fault conditions. In Section 4.3, data were acquired from a laboratory bearing testbed, and verification was conducted with three bearing faults. Last, in Section 4.4, the IEEE PHM data challenge dataset including run-to-failure ball bearing vibration data were used (Nectoux *et al*. 2012). In this section, an early detection study was conducted to detect the point where the fault gets started.

4.1 Case study 1: CWRU bearing dataset

4.1.1 Experimental setup

The CWRU bearing testbed consists of a motor, torque transducer/encoder, dynamometer, control electronics, and test ball bearing support motor shaft. To simulate bearing faults, bearing faults were seeded using electro-discharge machining. The size and type of faults used for verification are specified in Table 1. The vibration signal was collected using an accelerometer attached to the drive end, with a sampling rate of 48 kHz. Experiments were conducted in four operating environments with different RPM and motor loads.

4.1.2 Results

Fig. 5 shows the resulting anomaly scores for normal data and faults data. Normal data anomaly scores are represented by blue dots, a ball fault is represented by a red '*x*', an inner race fault by a yellow cross ('+'), and an outer race fault with a purple asterisk ('*'). The threshold, gray dotted line distinguishes fault data perfectly. The overall accuracy is shown in Table 2, and the proposed method was able to distinguish them all, regardless of the type and size of the faults.

4.2 Case study 2: PHMAP 2021 compressor dataset

4.2.1 Experimental setup

These data were collected from an oil-injection screw compressor. The screw and motor were connected by a belt, the rotation ratio was 2:1, the motor rotated at 3600 rpm, and the screw rotated at 7200 rpm. The data were collected with a sampling rate of 10544 Hz and a vibration sensor was attached to each motor and screw to collect data via 2 channels. The data have one normal class and three faults classes. In addition to bearing faults, these data also include defect data on unbalance and belt looseness during assembly. Detailed defect types are indicated in Table 3. Data are obtained from the compressor system, and it consists of two parts: the duration of rapid change in energy in the process of compression, and the duration of relatively constant energy. The part where the energy is relatively constant was used for learning and testing.

4.2.2 Results

The threshold of 1.002 was set according to Eq. (5). The anomaly detection accuracy for each fault is shown in Table

Table 3 Seeded fault type and description of faults

Fault type	Description
Unbalance	Unbalance between centers of mass and axis
Belt-Looseness	Looseness of V-belt connecting between motor pully and screw pully
Belt-Looseness	High looseness of V-belt
Bearing fault	Removing grease of ball bearing on motor, which induces its wear-out

Table 4 Anomaly detection accuracy for each fault

Unbalance	Bearing	Looseness	Looseness (High)
100%	99.78%	99.89%	100%

4. It shows a more than 99% detection accuracy for all faults, and a high fault detection accuracy for unbalance and looseness defects in addition to bearing defects. These results demonstrate that the proposed algorithm is capable of detecting structural faults such as belt looseness and unbalance as well as bearing faults.

4.3 Case study 3: Laboratory testbed

4.3.1 Experiment setup

This section covers the experimental setup used to validate the proposed method. Fig. 6 shows the bearing testbed used to collect the vibration data and for the bearing seeded fault tests. Two vibration sensors PCB 352C34 were mounted on the axial point and horizontal point of the bearing housing A and the bearing housing B. Type 6205 steel NSK ball bearings were used for testing. Three fault types including ball fault, inner race fault, and outer race fault were simulated on the ball bearing. The ball fault, inner race fault, and outer race faults were generated by scratching the steel race surfaces with a diamond tip. The vibration data were measured from the artificial fault seeded ball bearing at a rated rotation speed of 3000 rpm for 120 seconds at a sampling rate of 25.6 kHz. For consistency, the vibration sensor was placed in the same location for all data acquisitions. To validate the proposed method, fault detection was performed on the ball, inner race, and outer race faults.

Table 5 Anomaly detection accuracy for laboratory bearing testbed

Ball fault	Inner race fault	Outer race fault
100%	100%	100%

4.3.2 Results

The results of abnormal detection on the bearing testbed are shown in Table 5. Three fault types including ball fault, inner race fault, and outer race fault were simulated on the ball bearing. Fault detection of 100% accuracy was possible for all bearing faults.

4.4 Case study 4: IEEE PHM 2012 data challenge dataset

4.4.1 Experimental setup

The experiment was conducted with Pronostia, a laboratory research platform, and the bearing degradation data were collected by performing an accelerated life test. This dataset is run-to-failure data, and the operation began in a normal state and was stopped when the magnitude of the vibration exceeded 20 g. Data were collected at a

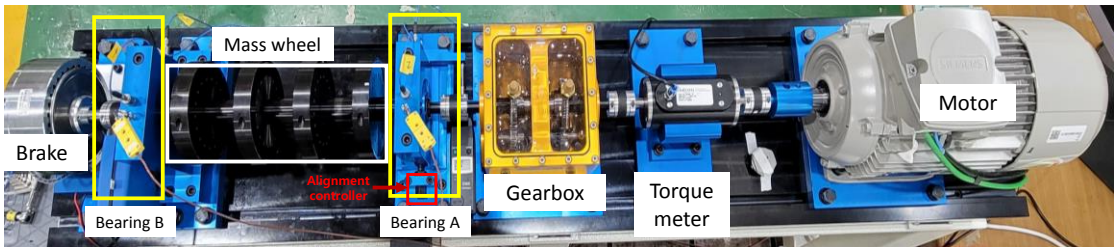
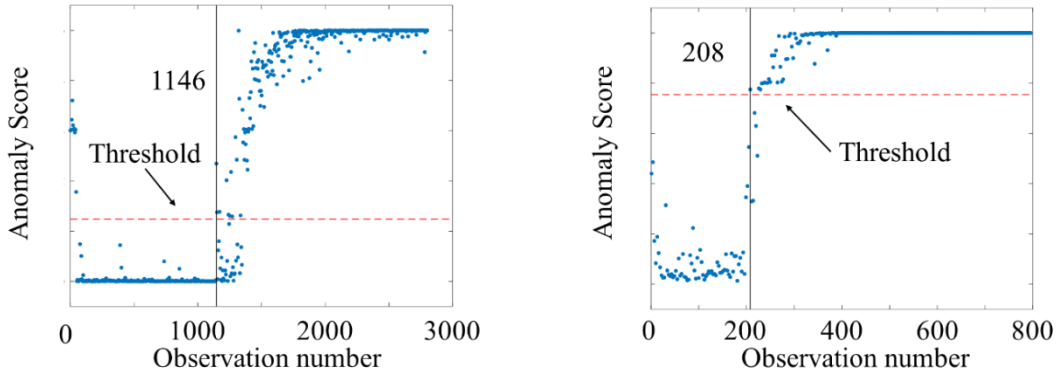


Fig. 6 Bearing testbed



(a) Bearing 1-1 (b) Bearing 2-2

Fig. 7 Anomaly score for Prognostia data

Table 6 Anomaly start point compared with the winning paper (1): proposed method, (2): winning paper, (3): total number of samples

	Predicted anomaly starting point (1)	True anomaly starting point (2)	Error [(1)-(2)]/(3)x100 (%)
Bearing 1-1	1146	1218	-2.56%
Bearing 2-2	208	181	3.38%

Table 7 Extracted features for learning

Feature	Formula	Feature	Formula
Maximum value	$f_1 = \max(X)$	Kurtosis	$f_9 = \frac{1}{N}\sum_{i=1}^{N} X_i^4$
Mean value	$f_2 = \frac{1}{N}\sum_{i=1}^{N} X_i$	Waveform indicator	$f_{10} = \frac{f_7}{f_6}$
Minimum value	$f_3 = \min(X)$	Pulse indicator	$f_{11} = \frac{f_1}{f_6}$
Standard deviation	$f_4 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - mean(X))}$	Kurtosis index	$f_{12} = \frac{f_9}{f_7}$
Peak to peak	$f_5 = \max(X) - \min(X)$	Peak Index	$f_{13} = \frac{f_1}{f_7}$
Mean Amplitude	$f_6 = \frac{1}{N}\sum_{i=1}^{N} \lvert X_i \rvert$	Square root amplitude	$f_{14} = \frac{1}{N}\sqrt{\sum_{i=1}^{N} \lvert X_i \rvert^2}$
Root mean square	$f_7 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} X_i^2}$	Margin indicator	$f_{15} = \frac{f_1}{f_{14}}$
Skewness	$f_8 = \frac{1}{N}\sum_{i=1}^{N} X_i^3$	Skewness indicator	$f_{16} = \frac{f_8}{f_7}$

sampling rate of 25.6 kHz. Verification was performed under two different operating conditions. Early detection was performed to detect the starting point of a defect, and compared with the winner's paper (Sutrisno *et al*. 2012). Since this dataset was not a labeled dataset, we assumed that the first part of the data was normal and used for training.

4.4.2 Results

After the training process using the initial part of the data, anomaly score for data collected from the start to the end of the experiment was calculated. Fig. 7 shows the anomaly score from start of operation until the machine stopped. Since the IEEE dataset is run-to-failure data, the state of the machine changes over time, so it cannot be assumed that normal data is being extracted from a gaussian distribution; so the threshold was set to the maximum value of the anomaly score of normal data. For Bearing 1-1, an abnormality starts to be detected at the 1146th point, and for Bearing 2-2, at the 208th point. Compared with the actual anomaly point mentioned in the winning paper, the proposed method predicted the anomaly starting point with small error.

4.5 Comparison with existing method

The proposed algorithm was compared with four existing methods, including two classical AD methods, the One-Class SVM (Scholkopf *et al*. 2001) and Local outlier factor (Breunig *et al*. 2000); and with two deep learning-based anomaly detection algorithms, the Deep-Support Vector Data Description (Ruff *et al.* 2018) and GAN based anomaly detection (Jiang *et al*. 2019). The input data used in the two classical AD methods and GAN-based anomaly detection algorithms used the features extracted from raw vibration data as input. The extracted features are shown in Table 7.

Table 8 AUROC results for five networks on fault type

PHMAP 2021 Data	OC-SVM	LOF	Jiang *et al.* (2019)	Deep-SVDD	Proposed Method
Bearing fault	0.970	0.917	0.949	0.899	**0.999**
Unbalance	0.879	0.949	0.850	0.863	**1.0**
Looseness	0.892	0.917	0.693	0.939	**0.999**
Average	0.907	0.929	0.83	0.9	**0.999**

The PHMAP2021 challenge dataset was used as the dataset for performance comparison, and the AUROC values for each fault were compared, as shown in Table 8. For the CWRU bearing dataset, all of the deep learning-based anomaly detection showed very high classification performance with values close to 1, but with the PHMAP 2021 data, the proposed algorithm showed better performance than the other algorithms.

Unlike the other algorithms, the proposed algorithm was also trained on the data generated using the variation method. Since the data generated through the variation method contains external noise while maintaining normal representation, it was able to learn normal features robust to external noise and the proposed method achieve better performance.

5. Conclusions

This paper proposed an anomaly detection method for detecting faults in a rotating machine using a CNN network and a normal dataset. A self-labeling method was applied to learn the representation of normal data using a CNN network. Self-labeling is a technique that generates a new dataset through data augmentation and at the same time assigns new labeling according to the applied augmentation method. An anomaly score function was defined to quantify the anomality of data, and a threshold was set, assuming the normal sample was a gaussian distribution.

Also, by optimizing hyperparameters the proposed method achieved high performance regardless of the type of machine. To validate the performance of the proposed method, four case studies were conducted. As a result, abnormality detection was possible with very high accuracy of 100 % for the CWRU and laboratory bearing testbed, and 99.67 % for an oil-injection screw compressor dataset. The potential for finding early detection points was also confirmed using a run-to-failure bearing dataset. The proposed method showed higher AUROC than the existing GAN, autoencoder based anomaly detection algorithm. This study is expected to be used to evaluate the durability of

rotating machines using only a normal dataset. Future work will focus on identifying fault types using normal data only.

Acknowledgments

This research was supported by UNDERGROUND CITY OF THE FUTURE program funded by the Ministry of Science and ICT.

References

Akcay, S., Atapour-Abarghouei, A. and Breckon, T.P. (2019), "GANomaly: Semi-supervised anomaly detection via adversarial training", *Asian Conference Computer Vision 2018*, Perth Western, Australia, December.

Amar, M., Gondal, I. and Wilson, C. (2015), "Vibration spectrum imaging: A novel bearing fault classification approach", *IEEE Trans. Ind. Electron.*, **62**(1), 494-502. https://doi.org/10.1109/TIE.2014.2327555.

Breunig, M.M., Kriegel, H.P., Ng, R.T. and Sander, J. (2000), "LOF: Identifying density-based local outliers", *Proceedings of the 2000 ACM Sigmod International Conference on Management of Data*, Dallas, USA, May.

Case Western Reserve University Bearing Data Center (2022), Case Western Reserve University, Cleveland, USA. https://engineering.case.edu/bearingdatacenter.

Choi, S., Akin, B., Rahimian, M.M. and Toliyat, H.A. (2012), "Performance-oriented electric motors diagnostics in modern energy conversion systems", *IEEE Trans. Ind. Electron.*, **59**(2), 1266-1277. https://doi.org/10.1109/TIE.2011.2158037.

da Silva, A.M., Povinelli, R.J. and Demerdash, N.A.O. (2008), "Induction machine broken bar and stator short-circuit fault diagnostics based on three-phase stator current envelopes", *IEEE Trans. Ind. Electron.*, **55**(3), 1310-1318. https://doi.org/10.1109/TIE.2007.909060.

Dai, X. and Gao, Z. (2013), "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis", *IEEE Trans. Ind. Inform.*, **9**(4), 2226-2238. https://doi.org/10.1109/TII.2013.2243743.

Data challenge on PHMAP 2021 (2021), Asian Pacific Conference of the Prognostics and Health Management Society 2021, Seoul, Korea. http://phmap.org/data-challenge.

Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014), "Rich feature hierarchies for accurate object detection and semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, June.

Golan, I. and El-Yaniv, R. (2018), "Deep anomaly detection using geometric transformations", *Proceeding of NeurIPS 2018*, Montréal, Canada, December.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014), "Generative adversarial networks", *Proceeding of NeurIPS 2014*, Montréal, Canada, December.

Guo, L., Chen, J. and Li, X. (2009), "Rolling bearing fault classification based on envelope spectrum and support vector machine", *J. Vib. Control*, **15**(9), 1349-1363. https://doi.org/10.1177/1077546308095224.

Hoang, D.T. and Kang, H.J. (2019), "Rolling element bearing fault diagnosis using convolutional neural network and vibration image", *Cognit. Syst. Res.*, **53**, 42-50. https://doi.org/10.1016/j.cogsys.2018.03.002.

Isermann, R. (1984), "Process fault detection based on modeling and estimation methods-A survey", *Automatica*, **20**(4), 387-404. https://doi.org/10.1016/0005-1098(84)90098-0.

Jia, F., Lei, Y., Lu, N. and Xing, S. (2018), "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization", *Mech. Syst. Signal Pr.*, **110**, 349-367. https://doi.org/10.1016/j.ymssp.2018.03.025.

Jiang, W., Hong, Y., Zhou, B., He, X. and Cheng, C. (2019), "A GAN-based anomaly detection approach for imbalanced industrial time series", *IEEE Access*, **7**, 143608-143619. https://doi.org/10.1109/ACCESS.2019.2944689.

Jing, L. and Tian, Y. (2020), "Self-supervised visual feature learning with deep neural networks: A survey", *IEEE Trans. Pattern Anal. Mach. Intell.*, **43**(11), 4037-4058. https://doi.org/10.1109/tpami.2020.2992393.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012), "ImageNet classification with deep convolutional neural networks", *Proceeding of NeurIPS 2012*, Tahoe, USA, December.

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), "Gradient-based learning applied to document recognition", *Proc. IEEE*, **86**(11), 2278-2323. https://doi.org/10.1109/5.726791.

Meng, Z., Guo, X., Pan, Z., Sun, D. and Liu, S. (2019), "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis", *IEEE Access*, **7**, 79510-79522. https://doi.org/10.1109/ACCESS.2019.2923417.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-morello, B., Zerhouni, N., Varnier, C., Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-morello, B., Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Morello, B., Zerhouni, N. and Varnier, C. (2012), "PRONOSTIA : An experimental platform for bearings accelerated degradation tests", *Proc. IEEE Int. Conf. Prog. Health Manage.*, Denver, USA, June.

Noroozi, M. and Favaro, P. (2016), "Unsupervised learning of visual representations by solving jigsaw puzzles", *Proceeding of European Conference on Computer Vision*, Amsterdam, Netherlands, October.

Oh, H., Jung, J.H., Jeon, B.C. and Youn, B.D. (2018), "Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis", *IEEE Trans. Ind. Electron.*, **65**(4), 3539-3549. https://doi.org/10.1109/TIE.2017.2752151.

Randall, R.B. (2010), *Vibration-based Condition Monitoring*, John Wiley & Sons, Ltd , West Sussex, UK.

Ren, S., He, K., Girshick, R. and Sun, J. (2017), "Faster R-CNN: Towards real-time object detection with region proposal networks", *Proceeding of NeurIPS 2014*, Montréal, Canada, December.

Ruff, L., Vandermeulen, R.A., Görnitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E. and Kloft, M. (2018), "Deep one-class classification", *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, July.

Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U. abd Langs, G. (2017), "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery", arXiv preprint arXiv:1703.05921.

Scholkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001), "Estimating the support of a high-dimensional distribution", *Neur. Comput.*, **13**, 1443-1471. https://doi.org/10.1162/089976601750264965.

Shelhamer, E., Long, J. and Darrell, T. (2017), "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, June.

Srividya, A., Verma, A.K. and Sreejith, B. (2009), "Automated diagnosis of rolling element bearing defects using time-domain

features and neural networks", *Int. J. Min., Reclam. Environ.*, **23**(3), 206-215. https://doi.org/10.1080/17480930902916437.

Sutrisno, E., Oh, H., Vasan, A.S.S. and Pecht, M. (2012), "Estimation of remaining useful life of ball bearings using data driven methodologies", *2012 IEEE Conference on Prognostics and Health Management*, June.

Tao, S., Zhang, T., Yang, J., Wang, X. and Lu, W. (2015), "Bearing fault diagnosis method based on stacked autoencoder and softmax regression", *2015 34th Chinese Control Conference*, 6331-6335. https://doi.org/10.1109/ChiCC.2015.7260634.

Tian, J., Azarian, M.H. and Pecht, M. (2015), "Rolling element bearing fault detection using density-based clustering", *2014 International Conference on Prognostics and Health Management*, June.

Wang, H., Li, S., Song, L. and Cui, L. (2019), "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals", *Comput. Indus.*, **105**, 182-190. https://doi.org/10.1016/j.compind.2018.12.013.

Yin, S., Ding, S.X., Xie, X. and Luo, H. (2014), "A review on basic data-driven approaches for industrial process monitoring", *IEEE Trans. Ind. Electron.*, **61**(11), 6418-6428. https://doi.org/10.1109/TIE.2014.2301773.

Yoo, Y. and Baek, J.G. (2018), "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network", *Appl. Sci.*, **8**(7), 1102. https://doi.org/10.3390/app8071102.

Zeiler, M.D. and Fergus, R. (2014), "Visualizing and understanding convolutional networks", *Proceeding of European Conference on Computer Vision*, Zurich, Switzerland, September.

Zhang, R., Isola, P. and Efros, A.A. (2016), "Colorful Image Colorization", *Proceeding of European Conference on Computer Vision*, Amsterdam, Netherlands, October.