International Journal of Internet, Broadcasting and Communication Vol.16 No.3 121-138 (2024)
http://dx.doi.org/10.7236/IJIBC.2024.16.3.121

IJIBC 24-3-12

Fault diagnosis of linear transfer robot using XAI

Taekyung Kim, Arum Park*

Professor, Department of Big Data Analytics, Kyung Hee University
26, Kyungheedae-ro, Dongdaemun-gu, Seoul, Korea
tk_kim@khu.ac.kr

**Professor, Department of AI Service Marketing, Seoul Cyber University*
60, Solmae-ro 49-gil, Gangbuk-gu,Seoul, Korea
penellope007@iscu.ac.kr

Abstract

Artificial intelligence is crucial to manufacturing productivity. Understanding the difficulties in producing disruptions, especially in linear feed robot systems, is essential for efficient operations. These mechanical tools, essential for linear movements within systems, are prone to damage and degradation, especially in the LM guide, due to repetitive motions. We examine how explainable artificial intelligence (XAI) may diagnose wafer linear robot linear rail clearance and ball screw clearance anomalies. XAI helps diagnose problems and explain anomalies, enriching management and operational strategies. By interpreting the reasons for anomaly detection through visualizations such as Class Activation Maps (CAMs) using technologies like Grad-CAM, FG-CAM, and FFT-CAM, and comparing 1D-CNN with 2D-CNN, we illustrates the potential of XAI in enhancing diagnostic accuracy. The use of datasets from accelerometer and torque sensors in our experiments validates the high accuracy of the proposed method in binary and ternary classifications. This study exemplifies how XAI can elucidate deep learning models trained on industrial signals, offering a practical approach to understanding and applying AI in maintaining the integrity of critical components such as LM guides in linear feed robots.

Keywords: *Linear motion robot, Transfer Learning, Fault diagnosis, Linear rail misalignment, CNN, XAI*

1. Introduction

To enhance productivity, maintenance of production equipment is vital. Employing machine learning techniques to predict product malfunctions is increasingly common. However, from a management perspective, understanding the causes of failures and identifying necessary maintenance areas is essential. This understanding informs the development of new equipment, the creation of procedures for fault handling, and the establishment of governance systems.

The development and application of explainable artificial intelligence (XAI) is crucial in these respects. Yet,

Manuscript Received: May. 20, 2024 / Revised: May. 27, 2024 / Accepted: June. 1, 2024
Corresponding Author: penellope007@iscu.ac.kr
Tel: +82-02-944-5766
Professor, Department of AI Service Marketing, Seoul Cyber University, Korea

empirical evidence on how XAI can be beneficial is limited, raising the need for research in the application and management of smart factories utilizing the Internet of Things and AI. This research gap is particularly evident in the context of linear feed robots, which are crucial in various transportation tasks in industrial machinery and automated production lines. These robots, operating under high load conditions and continuous friction, are prone to wear and tear on components like rails, bearings, and blocks in linear feed guides. A failure in these components can lead to breakdowns in the entire system. Thus, the integration of XAI in diagnosing and monitoring the state of such components not only aids in pre-emptive maintenance but also provides crucial insights for future technological advancements and governance in smart manufacturing environments.

Linear feed robots are mechanical tools that require linear movement within a system. They are widely used in various transportation tasks requiring linear motion in industrial machinery, automated production lines, and industrial robots. In such environments, linear feed robots are employed for repetitive tasks under diverse conditions. High load conditions of transportation tasks, along with continuous friction in production lines and robots, lead to the aging and damage of rails, bearings, and blocks in linear feed guides. Therefore, diagnosing and monitoring the state of LM guides is crucial, as failures in the linear feed guide can lead to breakdowns of the entire system that incorporates them. Recently, various fault diagnosis methods based on data have been researched.

Data-based methods, aimed at obtaining generalized models, utilize large amounts of data to identify its characteristics and diagnose faults based on these findings. Non-learning-based methods such as KNN (K-Nearest Neighbors) [1] and PCA (Principal Component Analysis) [2], as well as learning-based methods including artificial neural networks [3], fuzzy logic [4], and support vector machines [5], can be used for data-based fault diagnosis. Furthermore, with the advancement of data accumulation, computational speed, and the development of smart manufacturing, various data-based fault diagnoses using deep learning are being studied.

Multilayer perceptrons [3], [6], sparse deep stacking networks [7], stack quality-based autoencoders [8], deep belief networks [9], long short-term memory networks [10], and CNNs [11]-[14] are primarily researched for fault diagnosis using deep learning. These deep learning models, with higher accuracy than traditional machine learning approaches like KNN and PCA, are being utilized in various industrial applications. Model-based or signal-based diagnostics offer the advantage of incorporating management ideas regarding specific issues, allowing for a strategic allocation of resources and capabilities for management. The interpretability of such models facilitates policy formulation, enhancing efficiency and effectiveness in institutional and managerial aspects. Compared to these advantages, deep learning, despite its superior performance, carries both positive and negative aspects due to its limited explainability. Therefore, integrating explainability into existing deep learning methods can offer significant benefits in managing and controlling production in smart factories. This integration allows for the interpretation of specific problems, aiding in strategic planning for resource and capacity management.

When a model's functioning can be explained, it becomes easier to establish policies that address institutional flaws or managerial inefficiencies. In this context, while deep learning's performance is commendable, its limited explainability presents a challenge. Balancing the positive aspects of deep learning's performance with the need for explainability is crucial, especially in the context of smart factory management and production control. Enhancing the explainability of deep learning models can significantly contribute to more efficient and effective decision-making processes in industrial settings.

Deep learning models, traditionally perceived as 'black boxes,' are now being unraveled through

advancements in explainable artificial intelligence (XAI). Recent research in XAI, specifically in deep learning contexts, has significantly mitigated the notion of these models as inscrutable [15, 16]. However, a substantial portion of XAI research is predominantly focused on image data, which inherently allows for more intuitive interpretation. Translating these XAI techniques, such as Gradient-Class Activation Map (Grad-CAM), to diagnostic models that utilize signal data poses a notable challenge. This difficulty underscores the necessity for dedicated XAI research tailored to signal data applications. Signal data, often derived from sensor networks within the Internet of Things (IoT) ecosystem, is a rich vein of real-time, analyzable information. This data type, ubiquitously sourced from mobile devices, smart city infrastructures, and various IoT deployments, provides an easily accessible and analyzable information stream. Considering the widespread availability and the critical role of signal data, it is imperative to direct XAI research towards better understanding and interpreting it. By aligning XAI development with signal data, we can more effectively tap into the expansive potential of IoT networks, enhancing applications across diverse sectors, including smart manufacturing, urban planning, and beyond.

In this study, we propose an XAI-based fault diagnosis method to detect clearance anomalies in wafer linear robots used in semiconductor equipment, applying the methods from references [17] and [18] that utilize Grad-CAM, FG-CAM, and FFT-CAM. For the detection of abnormal states, vibration and torque sensors are used, and the reasons for anomaly diagnosis in the model are interpreted by comparing 1D-CNN and 2D-CNN and visualizing the frequency band of Class Activation Maps (CAMs) using Grad-CAM, FG-CAM, and FFT-CAM. In conclusion, this study makes significant contributions in both engineering and management realms. From an engineering perspective, we have successfully proposed a novel method for detecting clearance anomalies in wafer linear robots used in semiconductor manufacturing. This method leverages the capabilities of explainable artificial intelligence (XAI), employing advanced techniques such as Grad-CAM, FG-CAM, and FFT-CAM for precise diagnosis. By utilizing vibration and torque sensors, our approach not only identifies abnormal states but also interprets the underlying reasons for these anomalies, offering a deeper understanding through the comparative analysis of 1D-CNN and 2D-CNN, and the visualization of Class Activation Maps (CAMs). From a management standpoint, our study serves as a pioneering example of applying XAI in smart manufacturing.

The insights derived from signal data-based XAI development and design provide valuable guidance for strategic decision-making in industrial settings. Our research illustrates how XAI can enrich management strategies, allowing for the formulation of more informed policies and practices. In doing so, it addresses a critical gap in the literature, demonstrating the practical application of XAI in managing and enhancing the operational efficiency of smart factories. Through this dual contribution, we believe our work not only advances the field of intelligent manufacturing systems but also lays the groundwork for future research aimed at integrating cutting-edge AI technologies into the realm of industrial management and operations. The structure of the paper is as follows: Chapter 2 describes the proposed fault diagnosis for linear feed robots, Chapter 3 discusses the composition of the dataset and experimental results, and the final Chapter 4 presents the conclusion.

2. Clearance Anomaly Diagnostic System for Linear Feed Robots

2.1 Structure of the XAI-Based Fault Diagnosis Method

The configuration of the clearance anomaly diagnostic system is as shown in Figure 1. The target one-axis feed robot is a wafer transport robot used in semiconductor equipment. First, vibration signals and torque signals are detected from the accelerometer and torque sensor. The torque signal has the advantage of being detectable through the torque sensor mounted on the equipment or the current sensor of the servo driver. The detected signals undergo data preprocessing before being used as input for the model. The input data is divided into 1D-CNN and 2D-CNN models, with the 1D-CNN model using raw signals in the time domain, and the 2D-CNN model generating a spectrogram that visualizes the power spectrum obtained through short-time Fourier Transformation (STFT). Spectrograms are widely used in visualizing audio signals [19]. The size of the spectrogram is adjusted to fit the input of the recognition algorithm. Subsequently, anomaly classification is performed using both one-dimension convolutional neural network (1D-CNN) and two-dimension convolutional neural network (2D-CNN) models. These models extract and learn different features and use this information to classify the presence and type of anomalies. Additionally, to visualize the classification results, the 2D-CNN model utilizes the Grad-CAM technique, while the 1D-CNN model explains the prediction results in the frequency domain using FFT-CAM and FG-CAM.

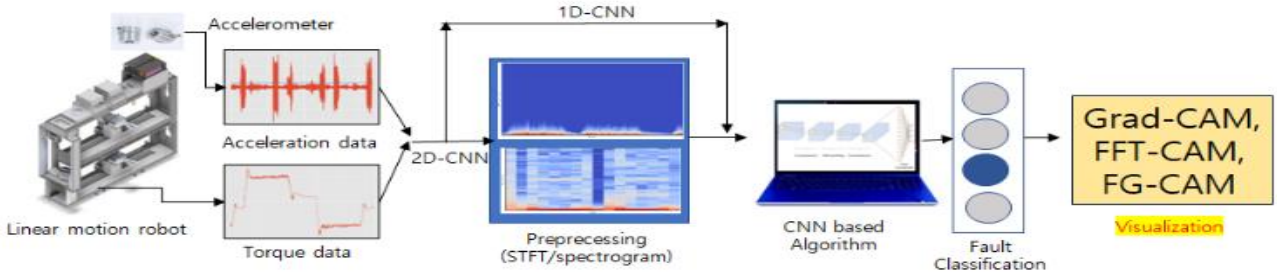


Figure 1. The proposed diagnostic system

(1) STFT

The STFT X(τ,w) of a signal x(t) is given by equation (1) as the Fourier transform of the product of x(t) and a window function w(τ), which is non-zero only over a short time interval. Window functions commonly used include the Hann window or the Gaussian window [20].

$$X(\tau, w) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t}\, dt \quad (1)$$

(2) Spectrogram

The spectrogram visualizes the values of $|X(\tau, w)|^2$ obtained from the STFT as a two-dimensional plot along the time and frequency axes, representing amplitude through color or brightness variations [19].

(3) 1D-CNN and 2D-CNN Models

The 1D-CNN is a neural network structure specialized for processing one-dimensional data and is widely used in the field of fault diagnosis utilizing deep learning. This model is particularly suitable for handling one-dimensional signals such as time series data and has the advantage of being able to process the data directly without additional transformation. Due to these characteristics, the 1D-CNN can be used to save computational effort even in cases requiring transformations to the frequency domain, such as Fourier or wavelet

transformations [17]. For signal data used in fault diagnosis, both torque and vibration signals are utilized. The length of the torque signal is set to 1,000 samples for the entire length, and the vibration signal is down-sampled by 1/10 to a length of 47,514 samples for use as model input. The 1D-CNN model uses the ReLU (Rectified Linear Unit) function as the activation function and transforms filter values by utilizing the maximum values in the kernel through the Max Pooling layer. Additionally, it performs classification tasks using the Global Average Pooling (GAP) layer to utilize the extracted features of filters. The experimented 1D-CNN model includes two 1D-CNN layers, each with a different number of filters. Specifically, 64 and 32 filters were used, and the kernel sizes were designed to be 100 and 50. These 1D-CNN layers, designed in this way, extract a wide range of signal characteristics and enable efficient data processing by reducing the size through Max Pooling. Before the output layer, there is one hidden layer, which processes the data using 100 filters. The constructed 1D-CNN model effectively analyzes the given signal data and performs discriminative tasks for fault diagnosis.

The 2D-CNN model, known as the VGG (Visual Geometry Group), is a deep learning model for image recognition capable of classifying the contents of images sized 224 × 224 into classes. The VGG-16 and VGG-19 models are used, each having 16 and 19 convolutional layers, respectively. In applying deep learning to the diagnosis of industrial systems, obtaining sufficient data to train the model can be challenging, especially with a scarcity of data on abnormal situations, leading to an imbalanced dataset. In such cases, transfer learning can be utilized. Transfer learning involves taking a model previously trained on a different problem with a different dataset, fixing some of its layers, and retraining it with data from a new domain [21]. This allows the model, initially trained for image classification, to be used for analyzing data in a different domain and detecting abnormal states, thereby helping to mitigate the issue of insufficient training data.

(4) Grad-CAM

One of the explainable artificial intelligence techniques, Class Activation Map (CAM) interprets neural networks by visually representing the location of pixels that influenced the result of a CNN [23]. This technique transforms the last convolutional layer of the CNN into a fully connected layer through global average pooling (GAP) and classifies images using the weights of each point in the fully connected layer as the weights of the activation layer. CAM provides visualization considering the location of pixels in the last convolutional layer of the CNN. To achieve this, the average value of each feature map is multiplied, and all the multiplied channels are summed up to create a heatmap. However, CAM has limitations as it can only be used in the last convolutional layer and may result in performance degradation due to the use of fully connected layers. Grad-CAM is one of the models developed to overcome these drawbacks of CAM, creating heatmaps using backpropagation gradients instead of weights [24]. The weights and formula for Grad-CAM are as follows. With this approach, Grad-CAM more effectively interprets the results of CNNs.

$$a_k^c = \frac{1}{h*w}\sum_{i=1}^{h}\sum_{j=1}^{w}\frac{\partial y^c}{\partial A_{i,j}^k} \tag{2}$$

$$Grad - CAM^c = ReLU\left(\sum_{1=1}^{N} a_k^c A^k\right) \tag{3}$$

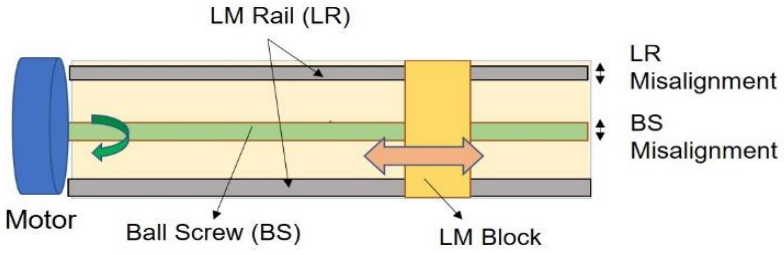


Figure 2. Definition of misalignments

(5) FG-CAM

For frequency band interpretation using the Grad-CAM technique, the ReLU function of Grad-CAM is substituted with Welch's method, a power spectrum transformation technique, to perform frequency analysis . The weights $\hat{a}_k^c$ of F^k in FG-CAM are as follows [22]:

$$\hat{a}_k^c = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial y^c}{\partial F_i^k} / \sum_{i=1}^{n} F_i^k \tag{4}$$

The calculated weights of the filters are used for frequency analysis by performing multiplication operations with each of the N filters in the output of the last 1D-CNN layer, and are then transformed into a power spectrum using Welch′s method.

$$FG - CAM^c = P_{\sum_{k=1}^{N}\hat{a}_k^c F^k}\,(\mathrm{f}) \tag{5}$$

(6) FFT-CAM

In FG-CAM, the calculated filter weights $\hat{a}_k^c$ are used with the power spectrum obtained by applying the direct Fast Fourier Transform (FFT) technique instead of Welch′s method to the result multiplied by filter F^k. Generally, the resolution of frequencies improves with more points in the signal to which FFT is applied. Therefore, it is advantageous to use the entire signal without segmenting it. FFT-CAM is calculated by the following formula [18]:

$$FFT - CAM^c = P_F\,(\sum_{k=1}^{N}\hat{a}_k^c F^k) \tag{6}$$

2.2 Dataset Composition

The dataset is composed of vibration and torque signals collected using accelerometers and torque sensors, as shown in Figure 2, from a robot driven at varying speeds in scenarios with clearance anomalies in the LM rail and ball screw shaft. Each dataset is categorized as shown in Table 1. The data is classified into five groups based on the state: normal condition, significant clearance anomaly in the linear rail, minor clearance anomaly in the linear rail, significant clearance anomaly in the ball screw shaft, and minor clearance anomaly in the ball screw shaft.

Table 1. Data class sets

Data set name	Meaning
N	Normal
LR-H	Large linear rail misalignment
LR-M	Minor linear rail misalignment

BS-H	Large ball screw misalignment
BS-M	Minor ball screw misalignment

3. Experiment and Results

3.1 Experimental Setup Configuration

The experimental setup, as shown in Figure 3, consists of a one-axis linear robot for wafer transportation equipped with an accelerometer and a torque sensor. The sampling frequency of the accelerometer is 48 kHz, and the torque sensor is 100 Hz.

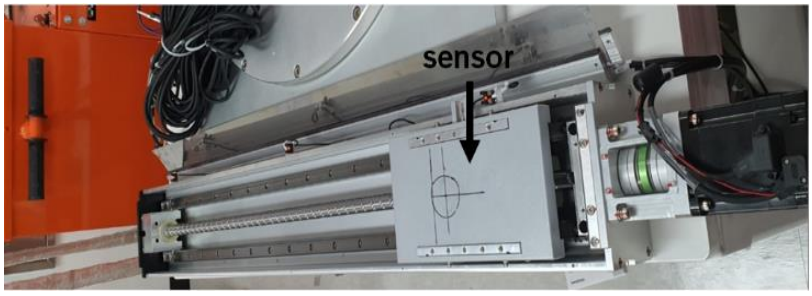


Figure 3. The LM rail sensor used in experiments

Figure 4 shows the spectrograms and original signals from the accelerometer under normal conditions, with rail clearance anomalies, and with ball screw shaft clearance anomalies. Figure 5 illustrates the spectrograms and original signals of the torque sensor under the same conditions.

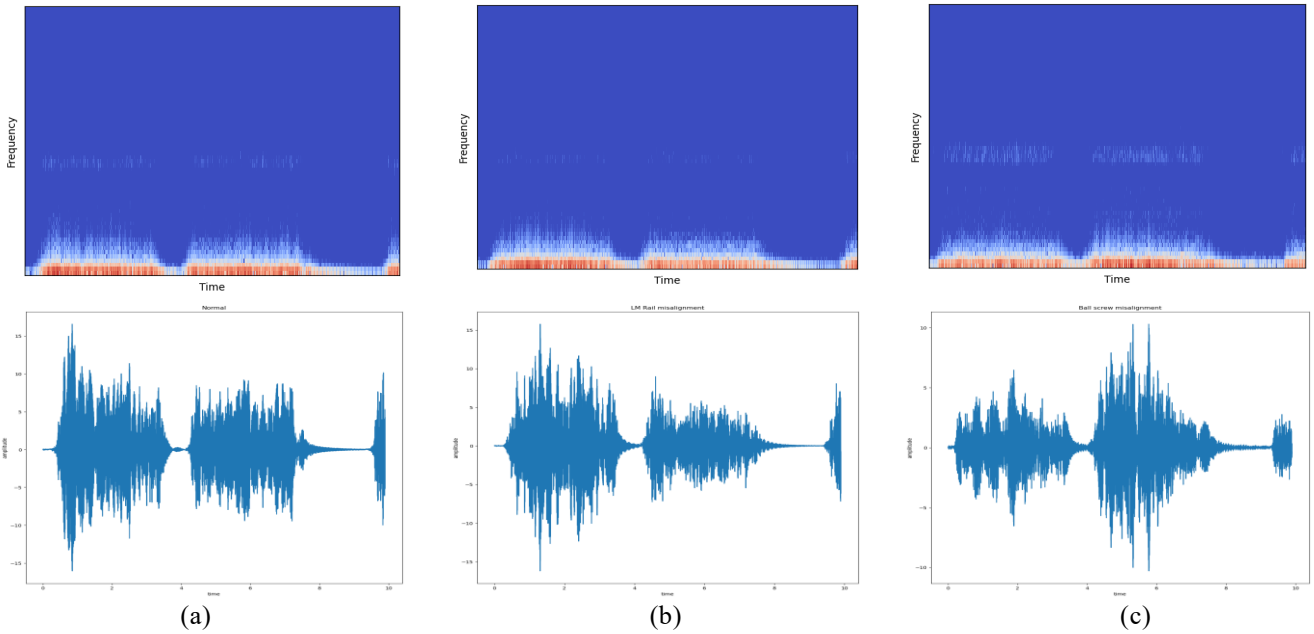
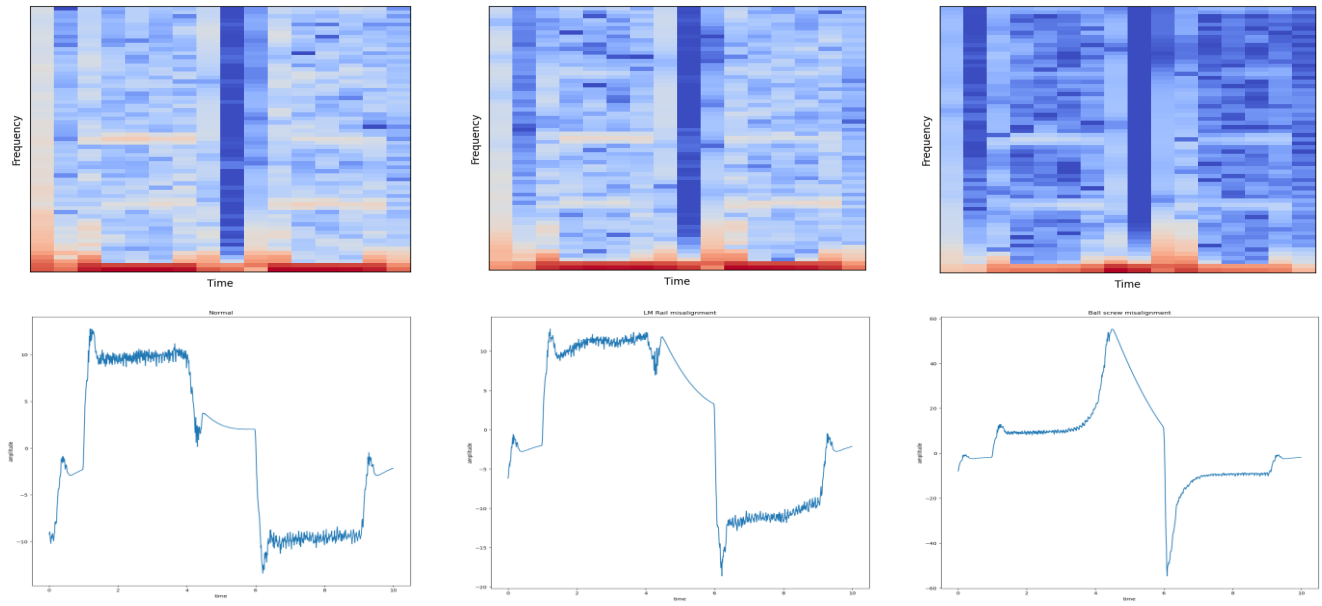


Figure 4. Acceleration signal spectrogram (above) and original signal (below)
(a) Normal (b) LM Rail misalignment (c) Ball screw misalignment



(a) (b) (c)

Figure 5. Torque signal spectrogram (above) and original signal (below)
(a) Normal (b) LM Rail misalignment (c) Ball screw misalignment

3.2 Training

The number of data used for training and testing is as shown in Table 2. The total displacement of the LM guide is 750 mm, and data was collected while operating at speeds ranging from 250 mm/s to 700 mm/s, divided into 10 stages. For each dataset, 100 sets of acceleration and torque signals were collected during operation. For model training, the Adam optimizer was used, and the loss function was categorical cross-entropy, with the number of repetitions (epochs) set to 150. In this case, the rail clearance anomalies were approximately 0.6 mm and 0.3 mm, and the ball screw clearance anomalies were 0.9 mm and 1.6 mm. For this experimental robot, the permissible rail clearance is 0.02 mm, and the ball screw clearance is 0.1 mm.

Table 2. Data sets for experiments

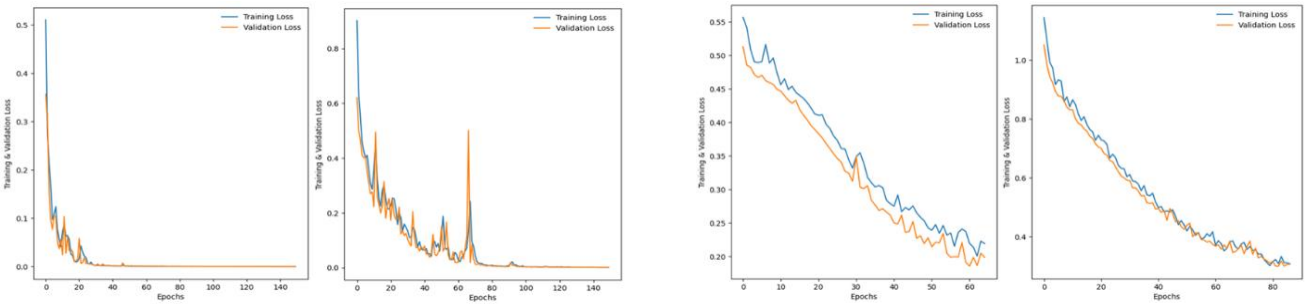
Data	# Training data	# Test data	Misalignment
N	80	20	-
LR-H	80	20	0.6 mm
LR-M	80	20	0.3 mm
BS-H	80	20	1.6 mm
BS-M	80	20	0.9 mm

The anomaly detection problem was divided into two types based on the number of classes to be distinguished, and each type was trained and tested separately. The loss function during the training process is

shown in Figure 6 for the accelerometer and as shown in Figure 7 for the torque sensor.

① 2 Classes: Normal/Abnormal

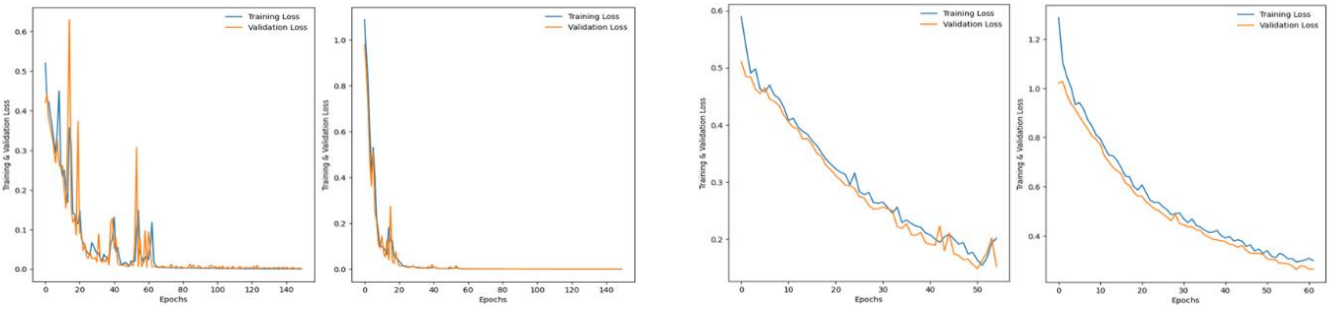
② 3 Classes: Normal/LM Rail Clearance Anomaly/Ball Screw Clearance Anomaly



(a) (b) (c) (d)

Figure 6. Loss function (acceleration data)

(a) 2 classes (1D-CNN) (b) 3 classes (1D-CNN) (c) 2 classes (2D-CNN) (d) 3 classes (2D-CNN)



(a) (b) (c) (d)

Figure 7. Loss function (torque data)

(a) 2 classes (1D-CNN) (b) 3 classes (1D-CNN) (c) 2 classes (2D-CNN) (d) 3 classes (2D-CNN)

3.3 Results

3.3.1 Anomaly Detection Accuracy

The test results of the implemented automatic diagnostic device are as shown in Table 3. When using the accelerometer, for the two-class problem of distinguishing normal from abnormal, accuracies of 100% and 97% were achieved depending on the model, while for the three-class problem of diagnosing the type of clearance anomaly, accuracies of 100% and 91% were observed.

Table 3. Diagnosis accuracy

Sensor Type	Model	Problem type	Accuracy %
Acceleration Sensor	1D-CNN	2 Classes	100
		3 Classes	100

	2D-CNN	2 Classes	97
		3 Classes	91
Torque Sensor	1D-CNN	2 Classes	100
		3 Classes	100
	2D-CNN	2 Classes	99
		3 Classes	98

In the case of using the torque sensor, for the two-class problem of distinguishing normal/abnormal, accuracies of 100% and 99% were achieved depending on the model, and for the three-class problem of diagnosing the type of clearance anomaly, accuracies of 100% and 98% were observed. In anomaly detection, using both vibration and torque sensors is deemed suitable for addressing clearance anomaly issues.

3.3.2 Confusion Matrix

Confusion matrices are used to evaluate the performance of classification models, visually representing the relationship between predicted and actual values [25]. The results of the final performance evaluation using test data are presented in confusion matrices. For the accelerometer, the confusion matrix graph is as shown in Figure 8, and for the torque sensor, the confusion matrix graph is as shown in Figure 9.

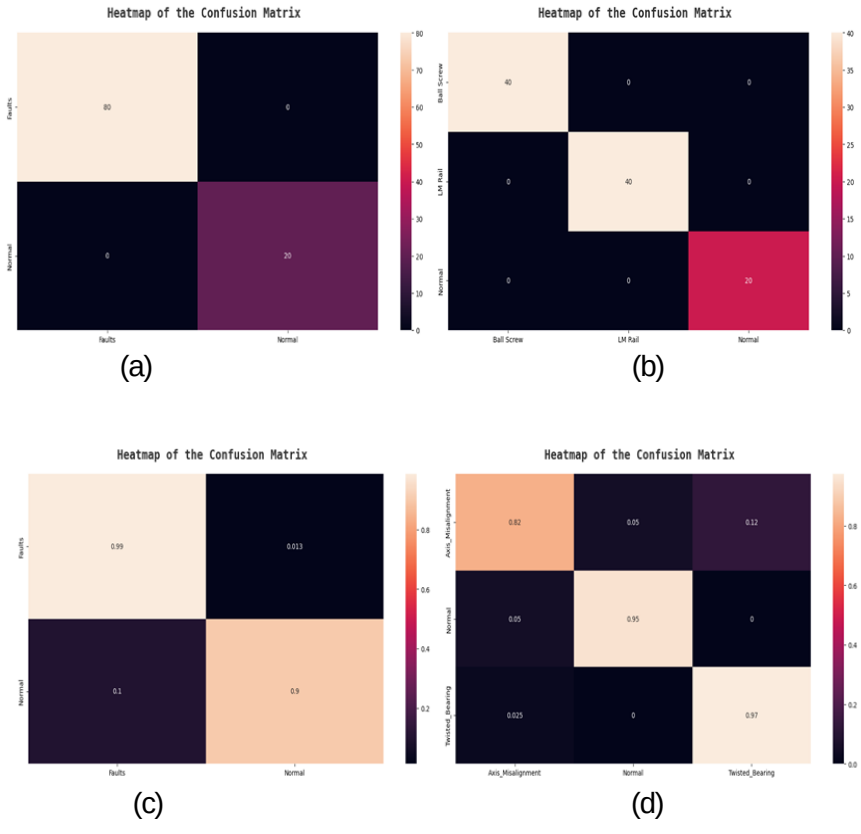


Figure 8. Confusion matrix graph of acceleration data case
(a) 2 classes (1D-CNN) (b) 3 classes (1D-CNN)
(c) 2 classes (2D-CNN) (d) 3 classes (2D-CNN)

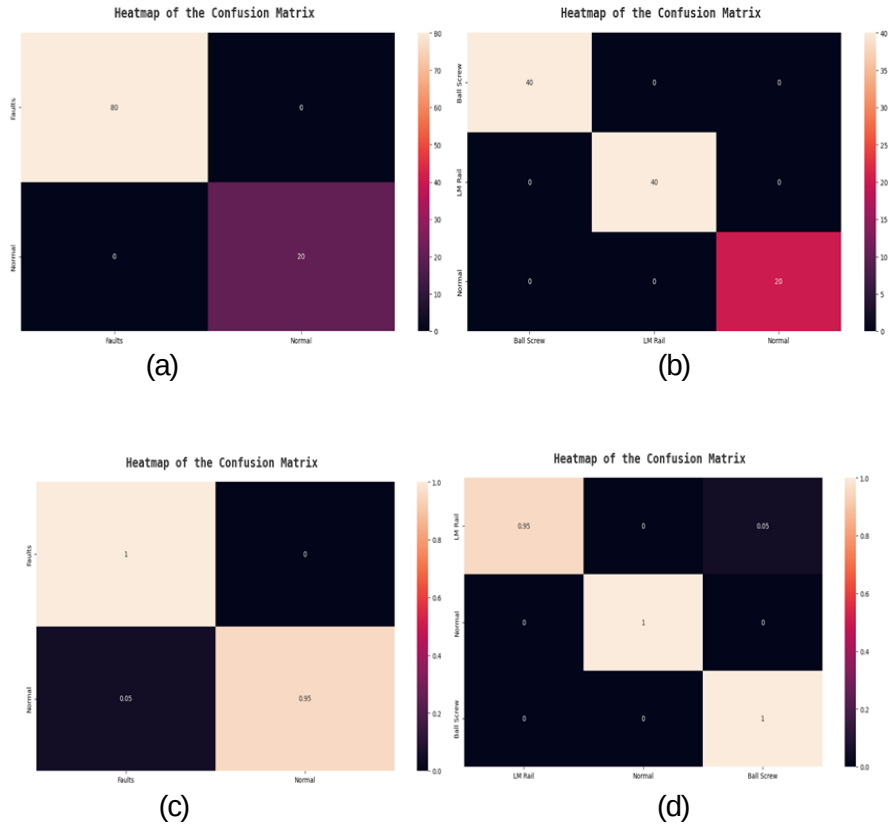


Figure 9. Confusion matrix graph of torque data case

(a) 2 classes (1D-CNN) (b) 3 classes (1D-CNN)

(c) 2 classes (2D-CNN) (d) 3 classes (2D-CNN)

3.3.3 Interpretation of XAI Models

(1) Applying Grad-CAM

Grad-CAM was utilized to analyze and explain the results of the CNN model. The results of applying Grad-CAM to the frequency images of operational state signals of linear feed robots for each sensor type are presented in Figure 10 and Figure 11.

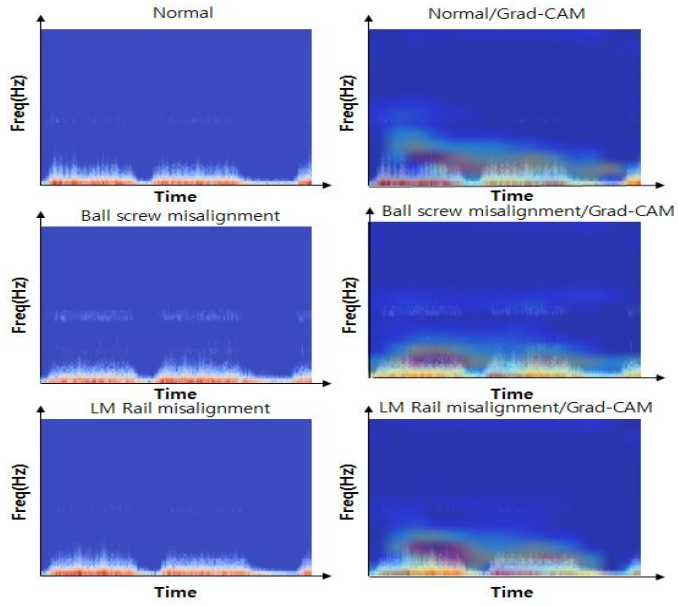


Figure 10. STFT and Grad-CAM result (acceleration data)

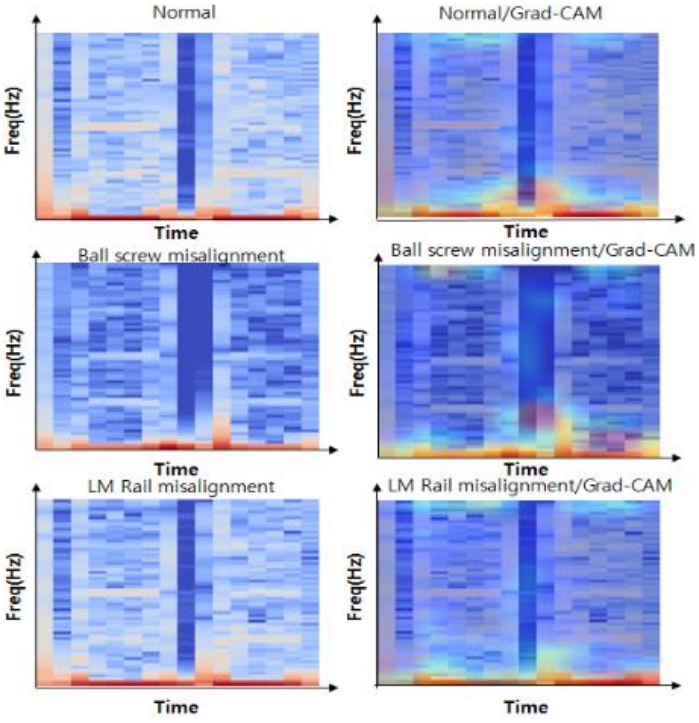


Figure 11. STFT and Grad-CAM result (torque data)

This revealed that in the cases of normal operation, ball screw misalignment, and rail misalignment, the frequency areas with higher weights in the model are located in the low-frequency band, and it was challenging to distinguish between them.

(2) Applying FG-CAM

For model interpretation, the power spectral density (PSD) graphs of the original signal's frequency band were compared with the PSD graphs of the Class Activation Map (CAM) using FG-CAM to derive results. This allowed for an interpretation of which frequency areas of the original signal in the frequency domain the model focuses on and utilizes for classification.

Figure 12 shows the average PSD graphs for the fault (ball screw side, rail side) and normal classes of the test data using an accelerometer. The blue box highlights areas where the frequency components of the original signal are lower compared to the class activation map. In the case of ball screw side faults, the primary frequency band referenced by the model in the original signal is approximately from 950Hz to less than 2,500Hz. For rail side faults, the model references weights in the frequency band from around 1000Hz, which is not prominent in the original signal, for classifying the rail side class. Additionally, for normal data, it can be observed that the model references the frequency band from 300Hz and from 800Hz to less than 2,500Hz with increased weight for classifying the normal class.

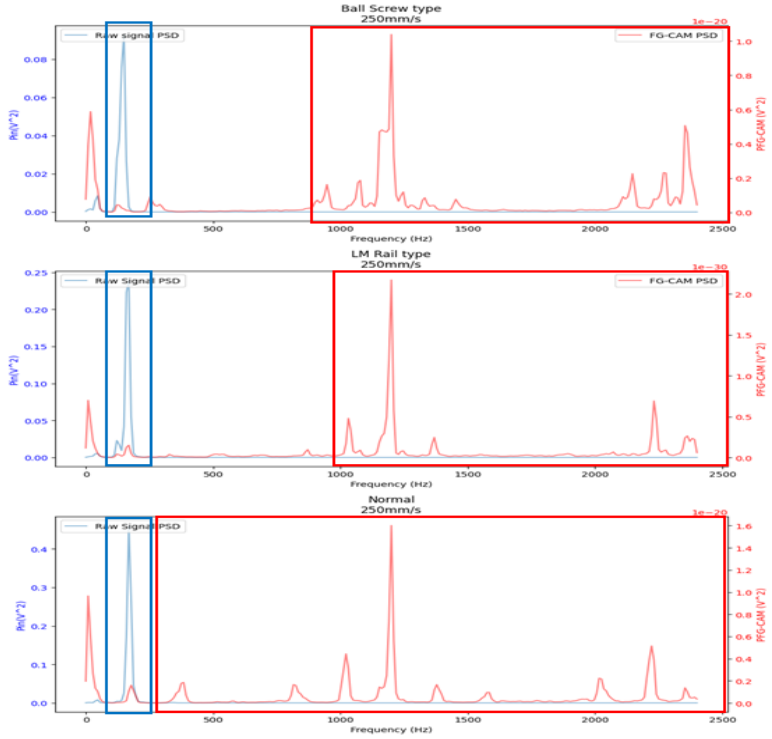


Figure 12. raw signal and FG-CAM result (acceleration data)

Figure 13 represents the average PSD graphs for the fault (ball screw side, rail side) and normal classes of the test data using a torque sensor. The blue box emphasizes areas where the frequency components of the original signal are lower compared to the class activation map. To classify normal data as normal, the model references the frequency band from 10Hz to less than 50Hz with increased weight, which is not prominent in the original signal. Unlike the normal state, it is challenging to distinguish frequency bands for each fault state. However, it can be observed that the model references and weighs the amplitude sections of the frequency

band differently for each type of fault, indicating amplitude differences between the fault types.

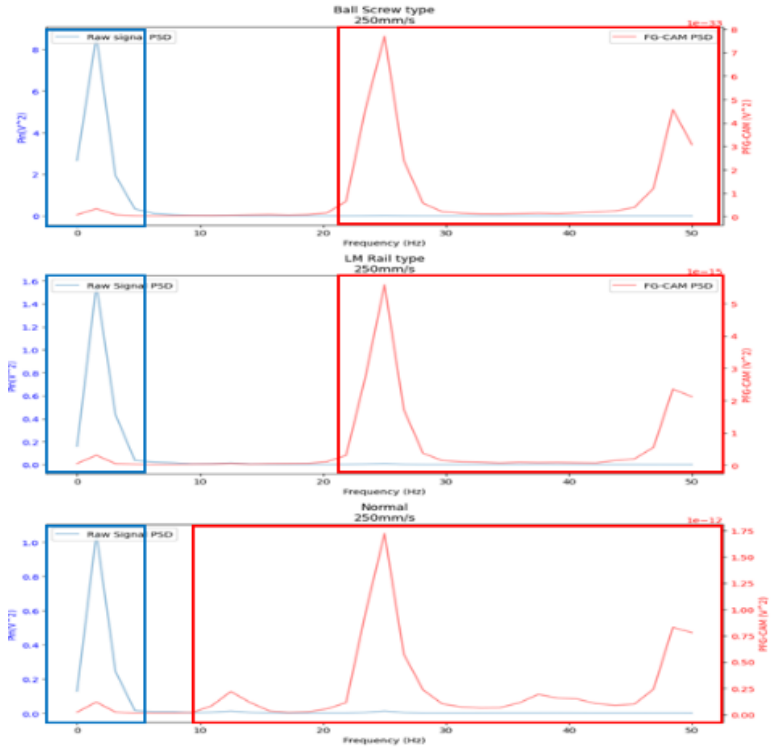


Figure 13. raw signal and FG-CAM result (torque data)

The results of the model interpretation experiment reveal that for all classes, including normal and abnormal states, the frequency bands primarily referenced by the model differ from the signal components in the original signal. This outcome indicates that the model assigns weights to the relevant information through learning for classification.

(3) Applying FFT-CAM

For interpreting the model results, model interpretation using FFT-CAM was conducted. The interpretation involves visually comparing the power spectrum density (PSD) graphs of the raw signal's frequency band used as input and the PSD graphs of the CAM signal obtained through the FFT-CAM technique. To explain the operation of the model, the frequency components of the signal contained in the actual raw signal and the frequency components of the characteristic signal focused on by the model are compared. This approach allows for an interpretation of which frequency bands of the raw signal the model selects and utilizes for classification. Figure 14 represents the case using an accelerometer and provides the average PSD graphs for the fault (ball screw side, rail side) and normal classes of the test data. The blue box indicates that the frequency component of the original signal falls in the low-frequency band around 200Hz. In the case of ball screw side faults, the model identifies frequency peaks at 950Hz, 1150Hz, 1250Hz, 2150Hz, 2300Hz, and 2400Hz in the original signal as contributing to classification. For rail side faults, the model references and weighs the 1250Hz frequency band in the original signal. Additionally, it can be observed that for normal data, the model strongly considers the frequency band from 1000Hz to less than 2500Hz for classifying it as normal.

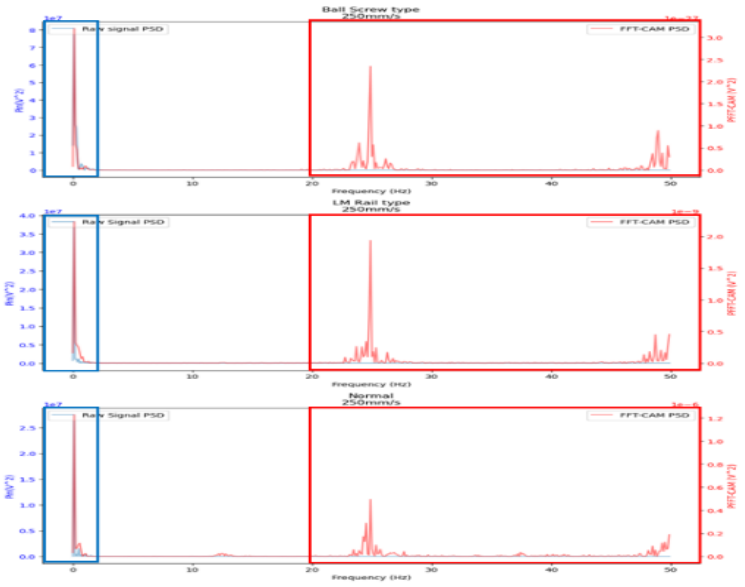


Figure 14. raw signal and FFT-CAM result (acceleration data)

Figure 15 addresses the case using a torque sensor and provides the average PSD graphs for the fault (ball screw side, rail side) and normal classes of the test data. The blue box indicates that the PSD graph of the frequency band of the original signal falls within the low-frequency band. The red box represents the CAM signal, which is not prominent in the PSD graph of the frequency band of the original signal. While it's challenging to distinguish which parts of the original signal's frequency band are identified for each fault (ball screw side, rail side) and normal data, significant differences in amplitude are observed. This suggests that the model assigns considerable weight to the amplitude sections of the frequency band.

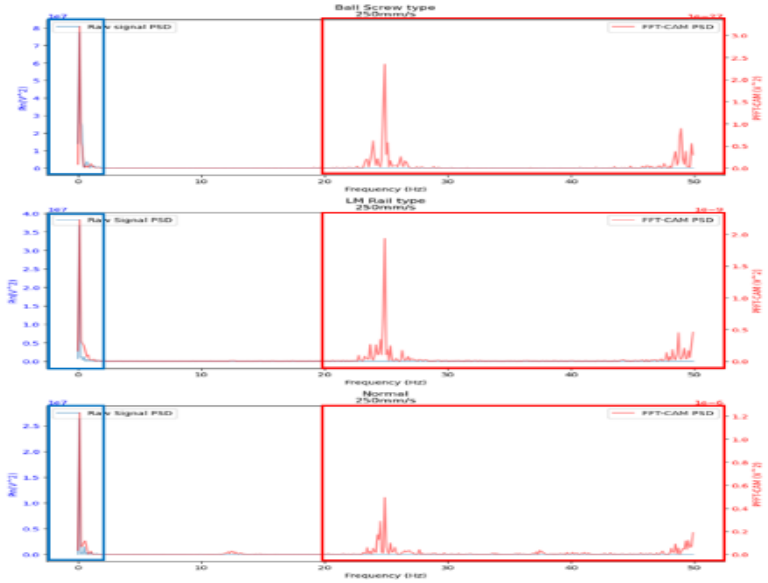


Figure 15. raw signal and FFT-CAM result (torque data)

The results of the model interpretation experiment reveal that for all classes, including normal and abnormal

states, the frequency bands primarily referenced by the model differ from the signal components in the original data. This outcome demonstrates that the model learns to assign weights to necessary information through learning for classification.

4. Conclusion

This study empirically verifies the role of AI features and chatbot system features in increasing the intention to use continuously in the context of chatbot services. Based on ECT, we have the direct and indirect effects of the three dimensions of AI-based features (i.e., personalization, personification, and social presence) and the two dimensions of the chatbot system (i.e., responsiveness and compatibility) on confirmation, satisfaction, and intentions to continuous use.

In this paper, an XAI-based fault diagnosis method using vibration and torque signals is proposed to diagnose clearance anomalies in one-axis robots for semiconductor equipment. The experimental results demonstrated high accuracy in both two-class and three-class classification problems using various sensors. When using the accelerometer, accuracies of 100% and 97% were achieved for the two-class problem of distinguishing normal/abnormal, and 100% and 91% for the three-class problem of diagnosing the type of clearance anomaly. With the torque sensor, accuracies of 100% and 99% were achieved for the two-class problem of distinguishing normal/abnormal, and 100% and 98% for the three-class problem of diagnosing the type of clearance anomaly. The use of both vibration and torque sensors was found to be effective in detecting clearance anomalies and their types.

Additionally, the reasons for the model's anomaly diagnosis were interpreted through the visualization of the frequency band of Class Activation Maps (CAM) using Grad-CAM, FG-CAM, and FFT-CAM. The results of applying the Grad-CAM technique showed that the frequency areas with higher weights were located in the low-frequency band, making it challenging to distinguish each. However, with FG-CAM and FFT-CAM, it was possible to identify the frequency bands of the original signal that the model focuses on and to detect the abnormal states.

As a result of this research, while interpreting the model with the Grad-CAM technique was challenging, the field of fault diagnosis requires further research and development. It is anticipated that future studies will improve the accuracy of fault diagnosis through better explainable artificial intelligence technologies and model enhancements. With such advancements, AI-based diagnostic methods can be expected to be applicable for anomaly detection in various industrial devices.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5A8080527)

References

[1] D. H. Pandya, S. H. Upadhyay and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl*, vol. 40, no. 11, pp. 4137–4145, Aug 2013. DOI:https://doi.org/10.1016/j.eswa.2013.01.033

[2] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park and I.-B. Lee, "Fault detection and identification of nonlinear

processes based on kernel PCA," *Chemometr. Intell. Lab*, vol. 75, no. 1, pp. 55–67, Jan 2005. DOI: https://doi.org/10.1016/j.chemolab.2004.05.001

[3] H. Hu, B. Tang, X. Gong, W. Wei and H. Wang, "Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks," *IEEE Trans. Ind. Informa.*, vol. 13, no. 4, pp. 2106–2116, Aug 2017. DOI: https://doi.org/10.1109/TII.2017.2683528

[4] X. Liu, L. Ma and J. Mathew, "Machinery fault diagnosis based on feature level fuzzy integral data fusion techniques," *presented at the IEEE Int. Conf. Ind. Informat*, Singapore, pp. 857–862, 2006, DOI: https://doi.org/10.1109/INDIN.2006.275689

[5] L. Ren, W. Lv, S. Jiang and Y. Xiao, "Fault diagnosis using a joint model based on sparse representation and SVM," *IEEE Trans. Instrum. Meas,* vol. 65, no. 10, pp. 2313–2320, Oct 2016. DOI: https://doi.org/10.1109/TIM.2016.2575318

[6] M. Heydarzadeh, S. H. Kia, M. Nourani, H. Henao and G. Capolino, "Gear fault diagnosis using discrete wavelet transform and deep neural networks," *presented at the 42nd Annu. Conf. IEEE Ind. Electron. Soc*, Florence, Italy, pp. 1494–1500, 2016. DOI: https://doi.org/10.1109/IECON.2016.7793549

[7] C. Sun, M. Ma, Z. Zhao and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informa.*, vol. 14, no. 7, pp. 3261–3270, Jul 2018. DOI: https://doi.org/10.1109/TII.2018.2819674

[8] X. Yuan, J. Zhou, B. Huang, Y. Wang, C. Yang and W. Gui, "Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy," *IEEE Trans. Ind. Informat*, vol. 16, no. 6, pp. 3721–3730, Jun 2020. DOI: https://doi.org/10.1109/TII.2019.2938890

[9] Y. Wang, Z. Pan, X. Yuan, C. Yang and W. Gui, "A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network," *ISA Trans*, vol. 96, pp. 457–467, Jan 2020. https://doi.org/10.1016/j.isatra.2019.07.001

[10] X. Yuan, L. Li and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Trans. Ind. Informat*, vol. 16, no. 5, pp. 3168–3176, May 2020. DOI: https://doi.org/10.1109/TII.2019.2902129

[11] S. Shao, S. McAleer, R. Yan,and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat*, vol. 15, no. 4, pp. 2446–2455, Apr 2019. DOI: https://doi.org/10.1109/TII.2018.2864759

[12] W. Sun, R. Zhao, R. Yan, S. Shao and X. Chen, "Convolutional discriminative feature learning for induction motor fault diagnosis," I*EEE Trans. Ind. Informat*, vol. 13, no. 3, pp. 1350–1359, Jun 2017. DOI: https://doi.org/10.1109/TII.2017.2672988

[13] R. Liu, G. Meng, B. Yang, C. Sun and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Informat*, vol. 13, no. 3, pp. 1310–1320, Jun 2017. DOI: https://doi.org/10.1109/TII.2016.2645238

[14] R. Liu, F. Wang, B. Yang and S. J. Qin, "Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions," *IEEE Trans. Ind. Informat*, vol. 16, no. 6, pp. 3797–3806, Jun 2020. DOI: https://doi.org/10.1109/TII.2019.2941868

[15] W. Samek, T. Wiegand and K.-R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," arXiv:1708.08296, 2017. DOI: https://doi.org/10.48550/arXiv.1708.08296

[16] D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Advanced Research Projects Agency*, DARPA/I20, 2017.

[17] M. S. Kim, J. P. Yun and P. G. Park, "An Explainable Convolutional Neural Network for Fault Diagnosis in Linear Motion Guide," *IEEE Trans. Ind. Informat*, vol. 17, no. 6, pp. 4036-4045, 2021. DOI: https://do

i.org/10.1109/TII.2020.3012989

[18] J. Ahn, “An Explainable 1D-CNN Deep Learning Method for Fault Diagnosis of Rotating Machinery Using FFT-CAM,” *Master's Thesis*, Graduate School of Industrial Engineering, Kyung Hee University, 2021.

[19] G. Manhertz and A. Bereczky, “STFT spectrogram based hybrid evaluation method for rotating machine transient vibration analysis,” *Mech. Syst. Signal Process*, vol. 154, 107583, 2021. DOI: (CrossRef Link)

[20] J. Han, S. Park and S. Hong, “Performance Evaluation of the Continuous Wavelet Transformation Data in Motor Fault Diagnosis through XAI Algorithm,” *Trans. Korean Inst. Electr. Eng*, vol. 71, no. 1, pp. 225-232, 2022.

[21] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv preprint arXiv:1409.1556, 2015. DOI: https://doi.org/10.1016/j.ymssp.2020.107583

[22] Y. Kim, H. Jeon and Y. K. Kim, “A Comparison Study of Ball Bearing Fault Diagnosis and Classification Analysis Using XAI Grad-CAM,” *Trans. Korean Inst. Electr. Eng*, vol. 71, no. 9, pp. 1315-1325, 2022. ISSN 1975-8359 [Print] / ISSN 2287-4364 [Online].

[23] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, “Learning deep features for discriminative localization,” in Proc. I*EEE Conf. Comput. Vis. Pattern Recognit*, pp. 2921–2929, 2016.

[24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis*, pp. 618–626, 2017.

[25] Truman, "What Is a Confusion Matrix?" Truman's Blog. [Online:https://truman.tistory.com/179]