

A Fault Prognostic System for the Logistics Rotational Equipment

Soo Hyung Kim[†] · Berdibayev Yergali · Hyeongki Jo · Kyu Ik Kim · Jin Suk Kim

Neoforce Co., Ltd.

물류 회전설비 고장예지 시스템

김수형[†] · 베르드바에브 예르갈리 · 조형기 · 김규익 · 김진석

주식회사 네오포스

In the era of the 4th Industrial Revolution, Logistic 4.0 using data-based technologies such as IoT, Bigdata, and AI is a keystone to logistics intelligence. In particular, the AI technology such as prognostics and health management for the maintenance of logistics facilities is being in the spotlight. In order to ensure the reliability of the facilities, Time-Based Maintenance (TBM) can be performed in every certain period of time, but this causes excessive maintenance costs and has limitations in preventing sudden failures and accidents. On the other hand, the predictive maintenance using AI fault diagnosis model can do not only overcome the limitation of TBM by automatically detecting abnormalities in logistics facilities, but also offer more advantages by predicting future failures and allowing proactive measures to ensure stable and reliable system management. In order to train and predict with AI machine learning model, data needs to be collected, processed, and analyzed. In this study, we have develop a system that utilizes an AI detection model that can detect abnormalities of logistics rotational equipment and diagnose their fault types. In the discussion, we will explain the entire experimental processes : experimental design, data collection procedure, signal processing methods, feature analysis methods, and the model development.

Keywords : Machine Learning, Anomaly Detection, Fault Diagnosis, Vibration Signal, Frequency Analysis

1. Introduction

Recently, digital transformation is underway due to logistics intelligence as Logistic 4.0 using data-based technologies such as IoT, Bigdata, and AI is accelerating. These core technologies of the 4th Industrial Revolution have become major industries throughout the operation and maintenance of logistics facilities due to their economic advantages [3].

With the recent activation of non-contact work due to changes in the social environment such as COVID-19, logistics centers are also being automated and unmanned. Therefore,

the use of Logistics 4.0 such as robotics, digital twin, IoT, and AI is being carried out in various forms. In particular, computational intelligence, or AI technology, is being attempted to optimize and maintain the logistics operations of automation facilities such as parcel sorter, transport facilities, and logistics warehouses.

However, despite the intensive growth of logistics intelligence, facility maintenance faces many difficulties. Logistics facilities are frequently disrupted by poor operating conditions such as high loads, and damages to their major parts are likely to cause major accidents such as the suspension of the entire logistics center. For instance, recently there was a suspension of Daejeon Hub Terminal, which accounts for one-third of total delivery volume of a South Korean logistics company that accounts for about half of the domestic delivery

volume. Then it caused delays in services and expensive costs to redistribute the holds, which were about 1.5 millions, to other terminals [5].

The main components of most logistics automation facilities are rotating mechanical components such as motors and bearings. The machines are typical cases of early detection as they always have potential elements of danger and large accidents.

In order to ensure the reliability of the facilities, Time-Based Maintenance (TBM) can be performed in every certain period of time, but this causes excessive maintenance costs and has limitations in preventing sudden failures and accidents.

Unlike TBM, the predictive maintenance using AI fault diagnosis model can automatically detect abnormalities in logistics facilities, predict future failures, and take proactive measures, so it can ensure stable and reliable system management at low maintenance costs. A Research has found that the predictive maintenance of mechanical facilities can reduce unexpected downtime losses by up to 20% and minimize maintenance costs by up to 10% [1].

The predictive maintenance market is expected to grow from 4 billion dollars in 2020 to 13 billion dollars by 2025. In particular, the delivery and logistics industry is expected to grow from 548 million dollars in 2020 to 1.7 billion dollars by 2025, and AI-based predictive maintenance technology is expected to be a core technology in the logistics industry in the future [4].

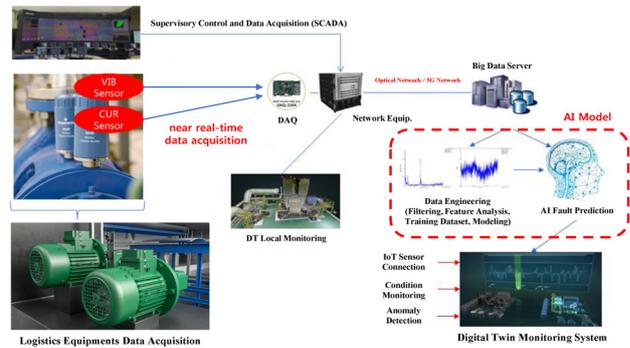
In this study, we created a fault diagnosis system for a rotating machine of a logistics center. Specifically, we collected normal and abnormal state of machines for training AI models, and we were able to detect failures of motor components such as bearing, rotor, and stator. In the rest of paper, I will explain how we design AI models, collect data, transform data, and train AI algorithms.

2. Model Design

2.1 Design of Fault Monitoring System

The facility diagnosis model in the fault monitoring system is for detecting abnormal signals from acquired sensor data and diagnosing which type of faults has occurred to the machine. The model becomes the main algorithm for the monitoring system, which collects vibration and current sensor data, processes them in order to diagnose any fault issue,

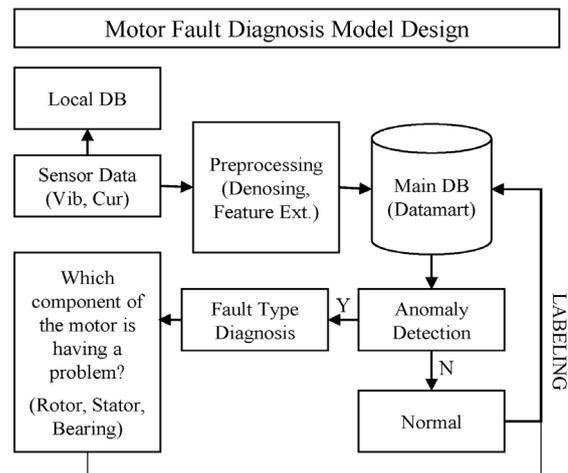
and displays a current state of the machine. The whole system design is shown in <Figure 1>. In this paper, we are going to focus on the AI model using vibration data as emphasized with a red-dotted line in the figure.



<Figure 1> Fault Diagnosis Monitoring System

2.2 Design of AI Fault Diagnosis Model

The algorithm of the fault diagnosis model is divided to two parts : anomaly detection and fault type diagnosis. The procedure of the AI fault diagnosis model is illustrated in <Figure 2> and is briefly defined as the followings.



<Figure 2> Motor Fault Diagnosis Model Design

- A data acquisition device collects a sensor data in real-time. This type of data is called time-waveform data, or raw data.
- We remove noise from the raw data by using a denoising method, and then calculate mathematically significant factors, so called feature parameters. These factors are

used for the anomaly detection. For fault type diagnosis, we convert the domain of the time-waveform data to the frequency domain and find the signature frequency for each fault types.

- The feature parameters are saved in a database and diagnosed by the anomaly detection. If the data is detected as an abnormal data, then it is further diagnosed to see which fault type it is.
- After the model diagnoses the condition of the machine, the data is defined as its condition and saved in the database.

3. Data Preparation and Transformation

3.1 Data Acquisition

We collected sensor data from an rotor kit, which is a motor-drive experimental setup as illustrated in <Figure 3>. An industrial conveyor belt is powered by a motor, which means that if we can predict a motor failure, we can prevent the malfunction of a conveyor belt. Therefore, we created the rotor kit in order to collect motor data and find fault signatures for the AI model.



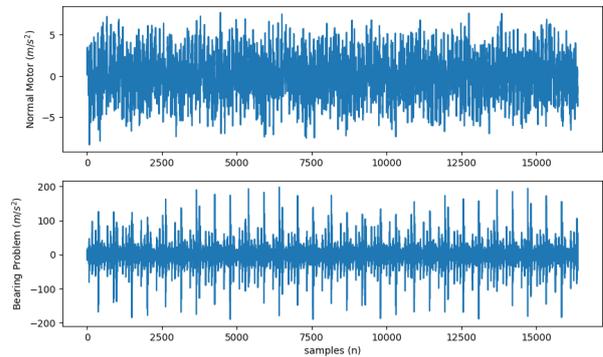
<Figure 3> Rotor Kit

In order to collect different types of fault, we prepared intentionally damaged motors. The specification of the motor used in the rotor kit are 1.5kW power supply and four poles. Although the motor is smaller than general motors for industrial conveyor belts, but we chose to use the size of a motor because of its price competitiveness. We damaged components inside the motor to create intentional failure of the motor, and since we were able to control the experiment, we labeled which part of the motor is broken. All fault labels and the number of data samples are listed in <Table 1>.

<Table 1> Fault Labels and Data Acquisition Amount

Component	Fault Type	Time (sec)
Bearing	Normal Bearing	1200
	Cracked Outer Race	300
	Cracked Inner Race	300
	Cracked Ball	300
	Cracked Cage	300
Rotor	Normal Rotor	600
	Broken Rotor Bar	300
	Eccentricity	300
Stator	Normal Stator	600
	Shorted Coil	300
	Single Phase	300

The data we achieved from the sensor is vibration data, measured in acceleration. The vibration is a sinusoidal waveform and is collected 16,384 samples per a second. The <Figure 4> illustrates the normal and abnormal vibration data we collected for the experiment.



<Figure 4> Normal and Abnormal Vibration Waveform

3.2 Noise Filtering

In order to analyze the vibration data more precisely, we need to remove noises from the sensor data. There are several noise removing methods such as autoregressive filtering, wavelet decomposition, etc. When removing noises from vibration signals, people often use the autoregressive filtering model, or so-called the AR model.

The AR model is a model that predicts a current state by summing up the past states regressively, <Equation 1>.

$$y_t = \sum_{i=1}^p \psi_i y_{t-i} + a_t \quad (1)$$

Since a discrete noise of the vibration is considered predict-

able due to its repetitive behavior, we can predict the noise using the AR model [9]. We can remove the noise from the original vibration in order to yield the denoised vibration signal as shown in <Equation 2>.

$$e(n) = x(n) - x_p(n), \quad (2)$$

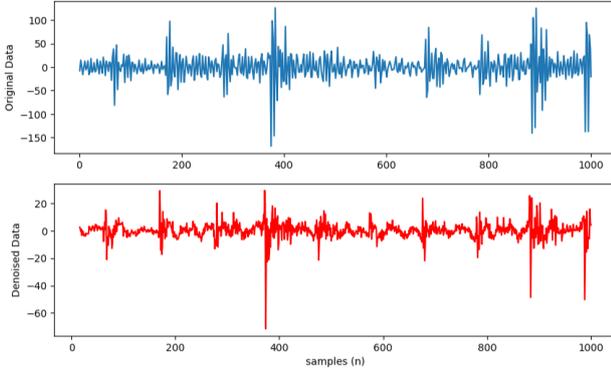
where $e(n)$ is the clean vibration, $x(n)$ is the original, and $x_p(n)$ is the discrete noise.

The discrete noise is obtained by <Equation 3>.

$$x_p(n) = - \sum_{k=1}^p a(k)x(n-k), \quad (3)$$

where n and k are time indexes, p is the order of the model, and $a(k)$ are regressive parameters.

After we applied the noise filtering method to the bearing data, we achieved a denoised fault data. The illustration in <Figure 5> describes the comparison between the original faulty signal and the denoised signal.



<Figure 5> The Comparison Between the Signals

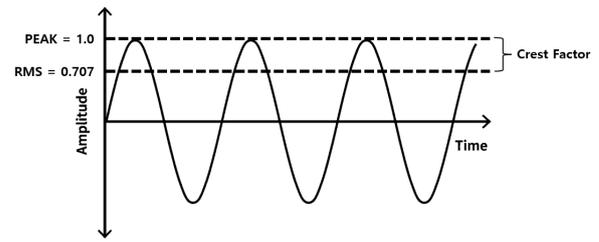
3.3 Feature Extraction for Anomaly Detection

The anomaly detection model needs to discriminate abnormality from time waveform data; therefore, we need to train the model with representative and summarized values, which is also called feature parameters. Feature parameters are inputs for training machine learning algorithms, and they are important factors to achieve high precision. We used three statistical indexes as feature parameters to demonstrate uniqueness of the vibration waveform: root mean squared, crest factor, and skewness. The equations for the indexes are shown in <Table 2>.

<Table 2> Time Waveform Feature Parameters

Statistical Indexes	Formula
Peak	$\max(x)$
Standard Deviation	$\sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}}$
Root Mean Square	$\sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$
Crest Factor	$\frac{PEAK}{RMS}$
Skewness	$\frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3}{STD^3}$

The root mean square and crest factor are key factors to represent time waveform [2]. Those factors are illustrated in <Figure 6>.



<Figure 6> Peak and RMS of Time Waveform

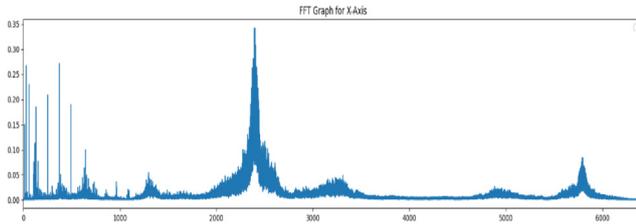
In addition to the two factors, skewness, which describes how sharp the peak is, can also show the uniqueness of vibration signals. Unlike the moderate normal vibration, the faulty vibration tends to have spikes as shown in <Figure 4>.

3.4 Feature Extraction for Fault Diagnosis

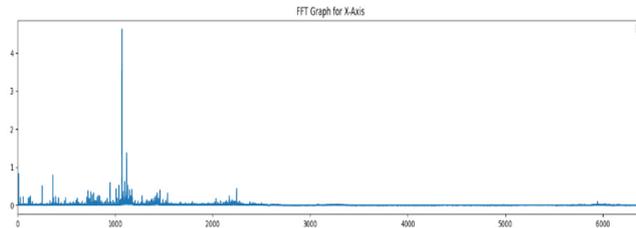
While statistically analyzing time waveform was enough for anomaly detection, we need to transform the time waveform data to a different type of waveform in order to find what type of fault the machine has. Each fault type has its specific frequency. That is, although they are rotating at the same time, a cracked rotor has different rotational occurrence to a cracked bearing because their sizes are different. Therefore, if we find the signature frequencies of each fault type, we can identify the corresponding failed components.

Therefore, we need to transform the time waveform data to the frequency waveform data. The most popular method is the Fourier Transform. We used a faster algorithm of the

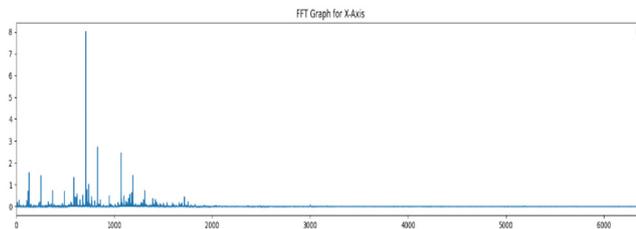
method to transform the time waveform data to the frequency data. The <Figure 7 A-C> illustrate the frequency data of a bearing, a rotor and a stator while they are abnormal [8].



<Figure 7A> Vibration of a Bearing in the Frequency Domain



<Figure 7B> Vibration of a Rotor in the Frequency Domain



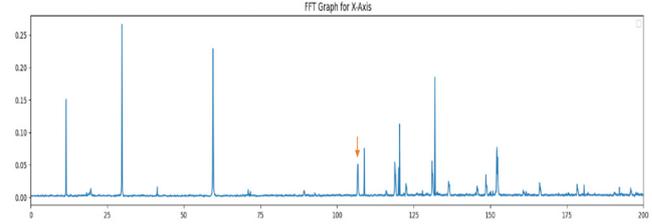
<Figure 7C> Vibration of a Stator in the Frequency Domain

To analyze the frequency data, we need to find the signature frequencies of the faults. These frequencies are defined as <Table 3>[2].

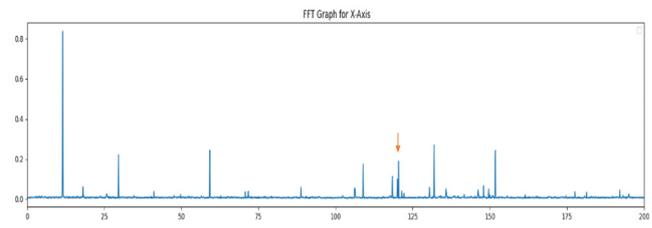
<Table 3> Vibration Signature Analysis for Motor Components

Component	Fault Type	Frequency Equation
Bearing	Cracked Inner Race	$BPMI = \frac{Nb}{2} S \left(1 + \frac{BD}{PD} \cos \beta \right)$
	Cracked Outer Race	$BPMO = \frac{Nb}{2} S \left(1 - \frac{BD}{PD} \cos \beta \right)$
	Cracked Ball	$BSF = \frac{Pb}{1Bd} S \left\{ 1 - \left(\frac{Bd}{Pd} \right)^2 (\cos \beta)^2 \right\}$
	Cracked Cage	$FTF = \frac{S}{2} \left(1 - \frac{BD}{PD} \cos \beta \right)$
Rotor	Broken Rotor Bar	$f_x = f_i (1 \pm 2ks)$
	Eccentricity	$f_x = f_i (1 \pm 2ks)$
Stator	Shorted Coil	$f_x = f_i \left\{ \frac{n}{p} (1 - s) \pm k \right\}$
	Single Phase	

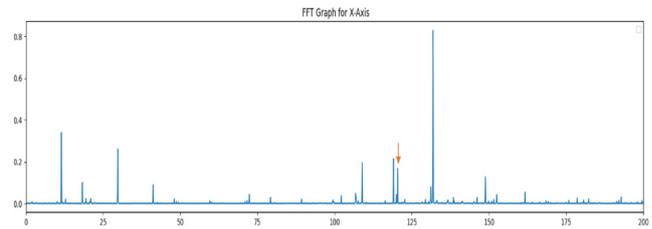
Using the frequency equations, we can find the signature frequencies from <Figure 7A-C>. The <Figure 8A-C> show the signature frequencies of each components.



<Figure 8A> Bearing – Outer Race Fault Frequency



<Figure 8B> Rotor – Eccentricity Fault Frequency



<Figure 8C> Stator – Shorted Coil Fault Frequency

As illustrated in <Figure 8A-C>, although the signature frequencies are not dominant, they are definitely depicted in the waveform. We record the magnitudes of the frequencies and use them as feature parameters for the fault type detection model.

4. Training Model

4.1 Training Dataset for Anomaly Detection Model

We calculated RMS, crest factor, and skewness from the time waveform vibration data and created a training dataset. Since we collected each 2,400 seconds of normal and abnormal data, we can create a dataset of 4,800 records of RMS, crest factor, and skewness. The dataset is divided into the training

dataset and testing dataset, respectively 7 to 3 ratio. The dataset for normal and abnormal data is shown in <Figure 9A-B>.

CH7_Crest	CH7_CrestFactor	CH7_Skewness	label
5.754406	0.415837	0.091572	0
6.450798	1.034777	0.055527	0
6.589952	0.710642	-0.005935	0
6.312672	0.659559	-0.032709	0
6.239767	0.631133	0.149524	0
...
4.657441	0.314608	-0.033291	0

<Figure 9A> Normal Training Dataset

CH7_Crest	CH7_CrestFactor	CH7_Skewness	label
5.772951	0.198278	-0.171975	1
4.557206	0.146160	-0.097673	1
5.194391	0.426468	-0.214161	1
6.157375	0.267198	-0.171244	1
4.767501	-0.017362	-0.138365	1
...
5.573474	0.270440	-0.306070	1

<Figure 9B> Abnormal Training Dataset

As illustrated in <Figure 9>, the normal data is labeled 0, and the abnormal data is labeled 1. Since there are only two types of data, we used a binary classification algorithm.

4.2 Anomaly Detection Model Algorithm

There are many binary classification methods such as the Decision Tree, the SVM, etc. In order to find the algorithm that has highest accuracy for the data, we used the AutoML library, so-called Pycaret. The library automatically validates the training dataset to all algorithms that it allows and evaluates the prediction score.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
qda	Quadratic Discriminant Analysis	0.9143	0.9469	0.9762	0.8733	0.9205	0.8286	0.8377	0.0130
et	Extra Trees Classifier	0.9071	0.9460	0.9286	0.8946	0.9090	0.8143	0.8193	0.0660
gbc	Gradient Boosting Classifier	0.9048	0.9398	0.9286	0.8903	0.9077	0.8095	0.8128	0.0230
rf	Random Forest Classifier	0.9024	0.9385	0.9333	0.8826	0.9058	0.8048	0.8091	0.0750
knn	K Neighbors Classifier	0.9000	0.9365	0.9476	0.8687	0.9053	0.8000	0.8057	0.0230
nb	Naive Bayes	0.9000	0.9254	0.9619	0.8616	0.9071	0.8000	0.8101	0.0130
xgboost	Extreme Gradient Boosting	0.8881	0.9324	0.8952	0.8881	0.8891	0.7762	0.7806	0.0250
lightgbm	Light Gradient Boosting Machine	0.8857	0.9424	0.8952	0.8837	0.8872	0.7714	0.7754	0.0170
ada	Ada Boost Classifier	0.8762	0.9282	0.8905	0.8710	0.8779	0.7524	0.7574	0.0230
dt	Decision Tree Classifier	0.8429	0.8429	0.8190	0.8639	0.8373	0.6857	0.6922	0.0130
lr	Logistic Regression	0.7167	0.7102	0.7619	0.6986	0.7261	0.4333	0.4400	0.0150
ridge	Ridge Classifier	0.7048	0.0000	0.7476	0.6897	0.7152	0.4095	0.4146	0.0110
lda	Linear Discriminant Analysis	0.7024	0.7082	0.7429	0.6880	0.7120	0.4048	0.4098	0.0130
svm	SVM - Linear Kernel	0.6643	0.0000	0.7429	0.5657	0.6368	0.3286	0.3679	0.0120
dummy	Dummy Classifier	0.5000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0120

<Figure 10> Anomaly Detection Training Result

We used 10-fold cross validation method to train the anomaly detection dataset. The cross validation method is for normalizing the prediction result after the model is trained. The <Figure 10> shows the prediction score for the anomaly detection.

The validation result shows that quadratic discriminant analysis predicts the most accurately with 91.4%. This is a result for self-testing with the training dataset; therefore, the model can perform less when it is evaluated with the testing dataset.

4.3 Training Dataset for Fault Diagnosis Model

For the fault type diagnosis model, we collected the signature frequencies and labeled each record with its fault type. The size of the dataset is 4,800 with normal state motor. Although it is a fault type detection model, we trained normal state in order to distinguish normal and fault states. The normal state is labeled as Motor. The <Figure 11> illustrates the dataset for fault type detection model.

BPFI	BPFO	BSF	FTF	BRB	ECC	SC	SP	label
0.017968	0.023670	1.0	0.041908	0.062258	0.032452	0.069380	0.358434	coll
0.012280	0.016338	1.0	0.020744	0.872466	0.013641	0.171178	0.207671	motor
0.067897	0.075890	1.0	0.065429	0.104157	0.119157	0.437410	0.625624	singlephase
0.093681	0.126608	1.0	0.143990	0.237446	0.064590	0.182293	0.179002	rotorbar
0.019210	0.035002	1.0	0.027380	0.459911	0.030543	0.093203	0.476183	inner
...
0.017722	0.044248	1.0	0.047238	0.533660	0.023209	0.036539	0.398780	inner
0.033300	0.023960	1.0	0.033102	0.447958	0.021154	0.093785	0.474027	inner
0.018708	0.055435	1.0	0.010606	0.208050	0.011219	0.095066	0.148962	cage
0.012274	0.022461	1.0	0.015319	0.832473	0.016852	0.215226	0.213175	motor
0.009784	0.023920	1.0	0.015572	0.821981	0.014259	0.175398	0.213827	motor

<Figure 11> Fault Diagnosis Model Dataset

The dataset for the fault type detection model is also divided, respectively 7 to 3 ratio.

4.4 Fault Diagnosis Model Algorithm

To effectively train the dataset for the model, we used four famous classification models, which are Extra Trees, Extreme Gradient Boost, Random Forest, and Catboost, with two classic models, Decision Tree and SVM [6]. The validation result is shown in <Figure 12>.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
rf	Random Forest Classifier	0.9838	0.9986	0.9838	0.9842	0.9833	0.9797	0.9801	0.8720
et	Extra Trees Classifier	0.9836	0.9986	0.9836	0.9850	0.9836	0.9795	0.9799	0.2380
catboost	CatBoost Classifier	0.9725	0.9995	0.9725	0.9759	0.9719	0.9656	0.9667	3.4480
xgboost	Extreme Gradient Boosting	0.9509	0.9989	0.9509	0.9608	0.9509	0.9386	0.9412	0.2300
dt	Decision Tree Classifier	0.9398	0.9622	0.9398	0.9459	0.9397	0.9245	0.9260	0.1960
svm	SVM - Linear Kernel	0.9348	0.0000	0.9348	0.9420	0.9349	0.9185	0.9203	0.2020

<Figure 12> Fault Diagnosis Training Result

5. Model Evaluation

We measured the performance of the models by calculating accuracy, precision, and recall. In addition to accuracy that is most definitely used as a metric to calculate the performance of a model, precision and recall are also important metrics to be considered because as the number of the prediction increases, the accuracy can be biased.

The followings are the equations for accuracy, precision, and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

For the Quadratic Discriminant Analysis classification algorithm for the anomaly detection model showed the performance as <Table 4>.

<Table 4> The Performance of Anomaly Detection

State	Accuracy	Precision	Recall
Normal	0.8864	0.8994	0.8965
Abnormal	0.8837	0.8945	0.8945

The anomaly detection model showed that it can find abnormality with 88% correctness. Despite the fact that this is a high accuracy, we can improve the accuracy by training more abnormal data.

The performance of the fault type detection model is shown in <Table 5>.

<Table 5> The Performance of Fault Type Detection

Model	Accuracy	Precision	Recall
RF	0.9810	0.9822	0.9920
ET	0.9834	0.9845	0.9986
CATB	0.9754	0.9762	0.9910
XGB	0.9491	0.9510	0.9989
DT	0.9285	0.9398	0.9622
SVM	0.9342	0.9348	-

We noticed that the performance of the fault type detection model is higher than the anomaly detection model. It is because we used more number of features for the fault type detection model. However, since the frequency analysis takes a longer

time than the time analysis, the fault type detection model costs higher than the anomaly detection model. Therefore, we can conclude that using both models in order to make them supplement to one another would yield the optimal result for the fault diagnosis monitoring system.

6. Concluding Remarks

In this paper, we developed a fault diagnosis model for rotating machines that are used in logistics facilities by analyzing vibration data from the rotor kit that imitates a motor in an industrial conveyor belt. As the result, we were successfully able to diagnose fault types of the data from the kit.

In logistics facilities, there are other important facilities. We can further extend this study by applying this fault detection algorithm to other facilities.

Acknowledgement

This work was supported by the Innopolis Foundation grant funded by the Ministry of Science and ICT (No. 1711175676, Development of a Fault Prognostic System for Power Plant Rotational Equipment based on AI Technologies)

References

- [1] Banks, M., 8 Causes of Lost Productivity in Manufacturing and How Managers Can Address Them in 2022, *Manufacturing Tomorrow*, March 14, 2022. <https://www.manufacturingtomorrow.com/story/2022/03/8-causes-of-lost-productivity-in-manufacturing-and-how-managers-can-address-them-in-2022/18384/>.
- [2] Berry, J., *Predictive Maintenance and Vibration Signature Analysis I*, first ed., Entek IRD International, 1993, pp. 2-6-6-64.
- [3] da Silva, R.M. Frederico, G.F., and Garza-Reyes, J.A., Logistics Service Providers and Industry 4.0: A Systematic Literature Review, *Logistics*, 2023, 7, 11. <https://doi.org/10.3390/logistics7010011>.
- [4] Global Market Trend Report - Prognostics and Health Management Market, *Innopolis Report*, 2021. pp. 1-17.
- [5] Jo, A.H., CJ Daehantongun, Daejeonteomineol Gadongjue Yeonmal Taekbae Bisang, *Newspim*, November 14, 2018. <https://www.newspim.com/news/view/20181113000806>.

- [6] Jung, H. and Kim, J.A., Machine Learning Approach for Mechanical Motor Fault Diagnosis, *Journal of Korean Society of Industrial and System Engineering*, 2017, pp. 57-64.
- [7] Kim, S., Berdibayev, Y., Park, J., Jung, J., Lee, S., Kim, K., Jo, H., Ahn, C., and Kim, J., A Fault Diagnostic Model for Electric Rotating Machines, *Proceedings of the 10th International Conference on Big Data Applications and Services*, 2022, Jeju Island, Korea, pp. 51-58.
- [8] Saruhan, H., Saridemir, S., Qicek, A., and Uygur, I., Vibration analysis of rolling element bearings defects. *Journal of Applied Research and Technology*, 2014, Vol. 12, No. 3, pp. 384-395.
- [9] Sim, J., Kim, S., Park, H.J., and Choi, J.-H. A Tutorial for Feature Engineering in the Prognostics and Health Management of Gears and Bearings, *Appl. Sci.*, 2020, Vol. 10, p. 5639. <https://doi.org/10.3390/app10165639>.
- [10] Wilson, S.J., Data Representation for Time Series Data Mining: Time Domain Approaches, *Wiley Interdisciplinary Reviews: Computational Statistics*, 2017, Vol. 9, No. 1, pp. e1392:1-6.
- [11] Yu, G.S. and Moon, Y.M., Development Trend of Smart Factory Facility Diagnosis and Predictive Maintenance Technology, *The Journal of The Korean Institute of Communication Sciences*, 2020, Vol. 37, No. 7, pp. 36-42.

ORCID

- Soo Hyung Kim | <http://orcid.org/0009-0007-1485-0312>
- Yergali Berdibayev | <http://orcid.org/0000-0001-7458-7009>
- Hyeongki Jo | <http://orcid.org/0009-0008-5365-1559>
- Kyu Ik Kim | <http://orcid.org/0009-0001-5251-2249>
- Jin Suk Kim | <http://orcid.org/0009-0000-0600-0552>