

Anomaly Detection and Fault Classification in Multivariate Time Series Using Multimodal Deep Models

Gunwoo Ryu¹, Nohyoon Seong²

^{1,2}*Aimtory, Seoul, 06249, Republic of Korea*

gunwoo.ryu@aimtory.com

nyseong@aimtory.com

ABSTRACT

In the realm of gear fault diagnosis, where various analytical methods often require extensive domain expertise, automation remains challenging due to diverse fault diagnosis tasks. To address these limitations, we propose a novel PHM algorithm integrating out-of-distribution detection and representation learning. Initial steps involve feature extraction using envelopes and fast Fourier transform (FFT). Representation Learning employs Transformers and Self-supervised learning for meaningful representations. The latent space values are then utilized for Out-of-Distribution Detection through kNN and classification, achieving a remarkable 99% accuracy. Our approach significantly enhances gear fault diagnosis automation, proving effective across diverse, unencountered problems.

1. INTRODUCTION

The realm of gear fault diagnosis, an area of enduring research interest, has seen the development of various analytical methods (Žvokelj, Zupan, and Ivan, 2016). Despite their efficacy, many of these established techniques require profound domain expertise and fail to facilitate the automation of gear fault diagnosis because of diverse fault diagnosis tasks (Lu, Wang, Qin, and Ma, 2017).

To overcome the limitations, we propose a novel PHM algorithm based on out-of-distribution detection and representation learning.

First, we conducted feature extraction using envelopes and the fast fourier transform (FFT). Subsequently, for Representation Learning, we employed Transformers and Self-supervised learning to obtain meaningful representations. Following this, we utilized the latent space values for both Out-of-Distribution Detection through kNN and classification. The results demonstrated an approximate

99% probability of identifying Out-of-Distribution data points. It was observed that the accuracy during Out-of-Distribution detection was significantly higher compared to cases where it was not performed.

In terms of practical contributions, our approach demonstrates effective handling of Out-of-Distribution cases, ensuring robust performance across a diverse array of problems that may not have been encountered during training.

2. PROPOSED METHOD

In this study, we address the proposed method of gear failure classification in the context of predictive maintenance for industrial machinery based on out-of-distribution detection.

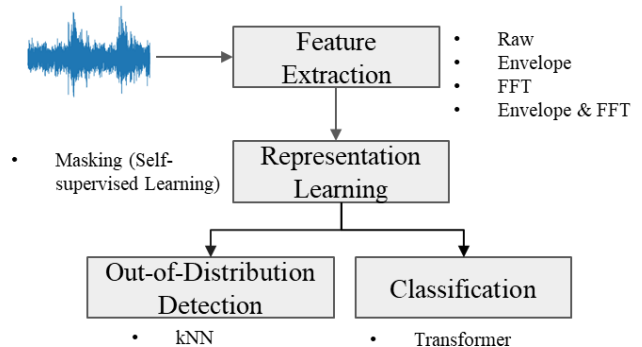


Figure 1 The Proposed Method

The approach encompasses a multi-step process integrating time series and frequency domain data analysis, representation learning, outlier detection, and classification.

2.1. Feature Extraction

We used two feature extraction techniques: envelope and FFT.

Since the length of each data is different, the data was divided into 1 second, and preprocessing was performed by combining FFT and Envelope Detection (magnitude of the

Ryu Gunwoo & NohYoon Seong. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

complex signal resulting from the Hilbert transform) for each axis.

2.2. Representation Learning

Convolutional Neural Networks (CNN) were applied to pre-processed data in both the time and frequency domains to extract features and patterns. Masked data was then fed into a transformer encoder to understand sequential data, as this forces the model to predict masked elements based on contextual information (Zerveas, Jayaraman, Patel, Bhamidipaty, and Eickhoff, 2021). The self-attention mechanism enhances the model's ability to capture complex dependencies in sequential data. A pretext task through a fully connected layer further refines the model's understanding of the data.

2.3. Out-of-Distribution Detection and Classification

After training the representation, we extracted the output of the last hidden layer and removed certain levels to generate a combination of trained OOD detectors. Each OOD detector used a K-nearest neighbor (KNN) approach (Sun, Ming, Zhu, and Li, 2022) to determine the OOD threshold based on the 75th percentile distance from the training data, where both the data in the distribution and the OOD data in the test set were within the distribution. We then removed the data determined to be OOD from all the generated detectors and used the KNN classifier to predict the defect level.

3. RESULT

3.1. Representation Learning

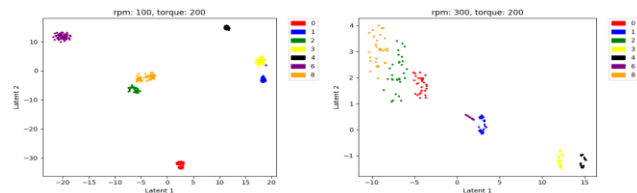


Figure 2 Representation Learning Result
The legend indicates the fault level

Figure 2 shows the extracted representation from the transformer encoder using t-SNE. Despite the fact that the model contains no information about torque and RPM, the data is effectively clustered by fault level.

3.2. Out-of-Distribution Detection

label	0	1	2	3	4	6	8	All
acc.	0.69	0.55	0.67	0.43	0.57	0.71	0.83	0.99

Table 1 Accuracy of each OOD detector

Table 1 shows the accuracy for the OOD detector obtained by removing the data for each level respectively in a ablation way. Each detector may perform slightly worse, but since the OOD decision is made by an ensemble of all detectors, the accuracy can be as high as 0.9994 with a high probability.

3.3. Classification

Accuracy	All	Without covariate	Time domain only	Frequency domain only
-	0.7826	0.7790	0.7646	0.7284

Table 2 Classification Accuracy

In the classification task, we divided the dataset into Training set, Validation set, and Test set with a ratio of 7:1:2. Ablation studies were conducted on the Training set and Validation set by removing one label at a time during experiments to consider the out-of-distribution problem. Subsequently, the experiments were repeated seven times, and the average accuracy of multi-class classification was reported in Table 2.

The overall accuracy is approximately 0.78, indicating a remarkably high performance. This figure is 0.12 higher than when out-of-distribution was not considered. Furthermore, considering static covariates (rpm, torque) resulted in higher accuracy compared to scenarios where they were excluded. This observation, coupled with the superior performance compared to considerations solely based on time domain or frequency domain, confirms the advantages of the proposed method.

4. CONCLUSION

In conclusion, our paper introduces a methodology for gear degradation analysis, emphasizing the need to consider multi-domain features and out-of-distribution issues. By combining advanced techniques such as CNN analysis, Transformer Encoding, and a customized classification approach, this methodology significantly advances our understanding of gear degradation.

ACKNOWLEDGEMENT

This research was supported by SW Convergence Cluster 2.0 through UIPA funded by the Ministry of Science and ICT (S0721-23-1011).

REFERENCES

Žvokelj, Matej, Zupan, S., & Prebil, I.. (2016). Eemd-based multiscale ica method for slewing bearing fault detection and diagnosis. *Journal of Sound & Vibration*, vol. 370, pp. 394-423. doi: 10.1016/j.jsv.2016.01.046

- Lu, C., Wang, Z.-Y., Qin, W.-L., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 130, 377–388. doi:10.1016/j.sigpro.2016.07.028
- Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021, August). A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 2114-2124).
- Sun, Y., Ming, Y., Zhu, X., & Li, Y. (2022, June). Out-of-distribution detection with deep nearest neighbors. In *International Conference on Machine Learning* (pp. 20827-20840). PMLR.

BIOGRAPHIES



Gunwoo Ryu is a data scientist of the A.I.mtory. He received a Master's degree in department of electronics and computer engineering and Bachelor's degree in department of electronic engineering from Hanyang University. His research interests include Signal Processing, Machine Learning.

NohYoon Seong is a CEO of the A.I.mtory. He received a Ph.D. degree in information systems from the College of Business, Korea Advanced Institute of Science and Technology (KAIST). He graduated from KAIST with a Bachelor degree in physics. His research interests include Signal Processing, Big Data Analytics, Machine Learning, Econophysics, and Applying Machine Learning and Econometric model in Business Analytics. He also participated in many industrial projects with big international companies as a data scientist.