Intelligent Bearing Fault Diagnosis Under Various Load Conditions Using Bias Mitigation

Seungyun Lee[1], Sungjong Kim[2], Minjae Kim[3], Heonjun Yoon[4], and Byeng D. Youn[5]

[1,2,3,5]*Department of Mechanical Engineering, Seoul National University, Seoul 08826, Republic of Korea*
lsy2140@snu.ac.kr
tjdwhd0423@snu.ac.kr
djm03254@snu.ac.kr
bdyoun@snu.ac.kr

[4]*School of Mechanical Engineering, Soongsil University, Seoul 06978, Republic of Korea*
heonjun@ssu.ac.kr

[5]*Institute of Advanced Machines and Design, Seoul National University, Seoul 08826, Republic of Korea*
bdyoun@snu.ac.kr

[5]*onepredict Inc., Seoul 06160, Republic of Korea*
bdyoun@snu.ac.kr

Abstract

Intelligent bearing fault diagnosis with domain adaptations has accomplished remarkable performances under various operating conditions. However, especially for different load conditions, the model bias due to the physical characteristics of bearing signals has not been considered. In the absence of handling bias, the root cause for generalization errors cannot be clarified under various load conditions. This paper thus demonstrates that certain bias exists in diagnostic models for different loads of bearings, and the main factor of bias is impulsiveness. The existence of bias is shown with quantitative analysis by applying fairness criteria to diagnostic models. Also, qualitative analysis is conducted with gradient-weighted class activation mapping (Grad-CAM) for vibration signals of bearings, which proves that the large amplitude of impulse can be the source of bias. To correct this impulsiveness bias, a framework of a fairness approach is newly proposed for bearing fault diagnosis under various loads. The process of correcting bias contains two steps: categorizing samples based on impulsiveness and training models with fairness criteria. Different from the previous domain adaptation-based approaches, the proposed method can achieve superior diagnostic performances by correcting bias that causes generalization errors. The effectiveness of the proposed method is validated with public-bearing datasets with various loads. The results show that the fairness approach can be the mainstream solution for fault diagnosis of rotary machines under different load conditions.

1. Introduction

Bearing fault diagnosis is crucial for the maintenance of mechanical systems. Bearings are essential components in rotating machinery, and their failure can lead to significant performance degradation and severe damage to the entire system (Ni, Q., Ji, J. C., Halkon, B., Feng, K., & Nandi, A. K., 2023). Early detection and diagnosis of bearing faults can prevent such issues, ensuring the reliability and longevity of machinery (Kumar, K., Shukla, S., & Singh, S. K., 2022). Consequently, various signal processing methods and fault diagnosis algorithms have been developed, with a growing emphasis on utilizing artificial intelligence techniques for more accurate fault detection.

In recent years, deep learning-based fault diagnosis methods have gained significant attention for their potential in real-world applications (Lee, J., Ko, J. U., Kim, T., Kim, Y. C., Jung, J. H., & Youn, B. D., 2024). These methods have demonstrated notable success, particularly with the implementation of transfer learning to diagnose bearing faults under various speed conditions. Transfer learning allows models to leverage knowledge from one domain and apply it to another, enhancing diagnostic performance even when training data is limited or differs from the testing conditions (Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C.,

& Liu, C., 2018). This approach has proven effective in adapting to the variable speed conditions commonly encountered in industrial environments.

While numerous studies have focused on diagnosing bearing faults under different speed conditions, research on diagnosing faults under varying load conditions remains sparse. Bearing signals exhibit distinct characteristics under different load conditions, unlike speed variations. This discrepancy poses challenges for applying transfer learning effectively, as the signal features that indicate faults can change significantly with varying loads. Therefore, existing transfer learning methods may not be suitable for all scenarios, necessitating the exploration of alternative approaches for load condition diagnosis.

This study aims to demonstrate the suitability of algorithm fairness in diagnosing bearing faults under different load conditions. By integrating the physical characteristics of bearings with deep learning models, we have developed a novel approach that successfully diagnoses faults across varying loads. This research highlights the importance of considering load variations in fault diagnosis and provides a robust framework for improving the reliability and applicability of deep learning-based methods in diverse operational conditions.

2. Theoretical Background

In this section, we will discuss the fundamental theoretical background related to the method proposed in this study. Specifically, we will explain the signals obtained when bearing fault in relation to fault frequencies. Additionally, we will delve into the concept of algorithmic fairness and its solutions.

2.1. Bearing fault signal

When a bearing fails, impacts occur between the bearing elements and the fault location, resulting in impulsive signals. In stationary conditions, these impulsive signals appear periodically and can be described using the shape factors of the bearing. Key fault frequencies include the Ball Pass Frequency Outer (BPFO), Ball Pass Frequency Inner (BPFI), and Ball Spin Frequency (BSF). These frequencies can be defined as follows (Randall, R. B., & Antoni, J., 2011):

BPFO:

$$BPFO = RPS \cdot \frac{n}{2}\left(1 - \frac{d}{D}\cos(\alpha)\right) \tag{1}$$

BPFI:

$$BPFI = RPS \cdot \frac{n}{2}\left(1 + \frac{d}{D}\cos(\alpha)\right) \tag{2}$$

BSF:

$$BSF = RPS \cdot \frac{D}{2d}\left(1 - \frac{d^2}{D^2}\cos^2(\alpha)\right) \tag{3}$$

where n is the number of rolling elements, RPS is the rotational speed, d is the diameter of the rolling elements, D is the pitch diameter, and α is the contact angle.

As the operating speed of the bearing changes, these fault frequencies also change. However, when the load on the bearing changes, the fault frequency remains the same, but the amplitude of the impulsive signals varies. This variation in amplitude under different load conditions presents a unique challenge for fault diagnosis algorithms, necessitating adaptive methods to accurately detect faults.

2.2. Algorithmic Fairness

Algorithmic fairness refers to the concept of ensuring that automated decisions made by algorithms are unbiased and equitable across different groups. This is crucial in applications where decisions can significantly impact individuals or groups, such as in healthcare, finance, and employment.

Fairness can be measured using several criteria: **Demographic Parity:** This criterion requires that the decision-making process result in equal positive outcome rates for all demographic groups. **Equalized Odds:** This criterion ensures that the algorithm has equal true positive and false positive rates across different groups. **Equalized Opportunity:** This criterion requires that the true positive rates are equal across different groups, focusing on providing equal opportunities for positive outcomes.

Solutions to achieve fairness in algorithms can be broadly categorized into three approaches: **Pre-processing:** Modifying the input data to remove bias before it is fed into the algorithm. This includes techniques such as re-weighting samples or transforming features to ensure neutrality. **In-processing:** Altering the algorithm itself to ensure fairness. This can involve incorporating fairness constraints into the optimization process during model training. **Post-processing:** Adjusting the output of the algorithm to achieve fairness. This can include re-ranking or modifying the decision thresholds for different groups. By integrating these fairness solutions, we aim to ensure that our fault diagnosis model performs equitably across various load conditions, providing reliable and unbiased results.

3. Proposed Method

This section illustrates the bias analysis procedure in bearing fault signals under different load conditions and the proposed algorithmic fairness approach for fault diagnosis. For the bias analysis, impulsiveness is calculated by the kurtosis values of the bearing fault signals. To mitigate this bias, fairness

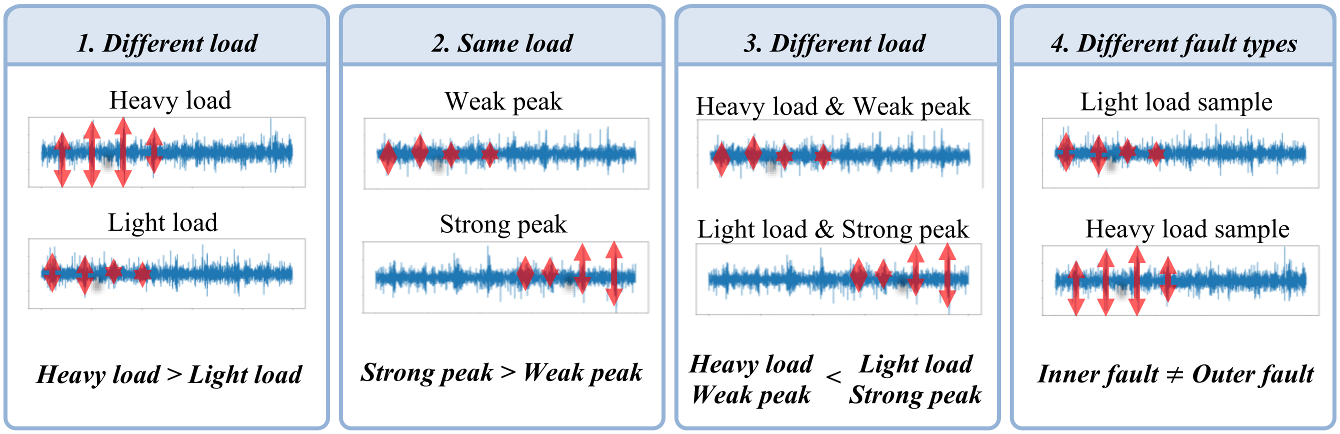


Figure. 1. Various scenarios for impulsiveness of the bearing signals under different load conditions.

training is applied by reweighting each instance of the fault signal considering the impulsiveness bias.

3.1. Bias Quantification

As described in Section 1, bearing fault signals under different load conditions can show specific characteristics in terms of impulsiveness. The impulsive signals show a large amplitude due to the strong impact between bearing components and defective areas. This degree of impulsiveness can affect data-driven models, which can be biased by different impulsiveness. Therefore, impulsiveness characteristics are quantified to evaluate the degree of possible bias in each instance.

A kurtosis feature is utilized to measure the impulsiveness in bearing signals. Kurtosis is a time-domain feature that can represent the impulsive component of the acceleration signal. This value can vary due to a lot of situations, as described in Figure. 1. Generally, an instance under a heavy load shows higher impulsiveness than under a light load. Under the same load conditions, there can be different impulsiveness due to the variance of the instance. In addition, there are a few cases that show a reverse trend: weaker impulsiveness under heavy load conditions than under light load conditions. Lastly, impulsiveness can be different due to the different health conditions: inner fault, outer fault, and ball fault.

To consider these characteristics of impulsiveness, the normalized kurtosis criterion technique is proposed. The proposed method can measure the degree of impulsiveness as a normalized value (i.e., zero to one). First, the bearing fault signal is divided into instances using the sliding window technique. For these instances, the kurtosis value for each fault type is calculated regardless of load condition. The calculated value can show multiple Gaussian distributions due to the different load conditions. To concatenate these distributions as normalized values, the cumulative distribution function (CDF) is calculated for all kurtosis values. Finally, each instance can have a normalized impulsiveness using the matched CDF values. For example, the high-impulsive instance shows the high value of kurtosis. Therefore, in the CDF, this instance is closer to one than the other instances. The CDF value can quantify the degree of impulsiveness for each instance.

3.2. Instance Reweighting

As described in Section 1, impacts occur between the bearing elements and the fault location, resulting in impulsive signals. The diagnosis model can be biased toward high-impulsive instances because they contain a lot of faulty information. To design a robust diagnosis model under different load conditions, the impulsiveness bias is mitigated using a reweighting method. By reweighting instances utilizing the quantified impulsiveness, the high impulsiveness is suppressed, and the low impulsiveness is highlighted in the model. The model is trained as follows:

$$L = -\frac{1}{N}\sum_{n=1}^{N} w_n y_n^{\text{true}} \log(y_n^{\text{pred}}) \tag{4}$$

where N is the number of samples in batch, w_n is the weight of each sample defined in Section 3.1, y_n^{true} is the true label (i.e., fault type of samples), and y_n^{pred} is the predicted label by the diagnosis model.

3.3. Model Training

A one-dimensional convolutional neural network (1D-CNN) is used as a model for bearing fault diagnosis. The 1D-CNN model has shown superior performance for extracting features from time-series data. For the model training, Eq. (1) is used for the loss function of the feature extractor and fault type classifier. To achieve accurate performance, the model is trained in 10 iterations using 10 random seeds, and the average performance is shown in Section 4. The overall flowchart is shown in Figure 2.

4. Experimental Validation

This section validates the proposed method by comparative studies and public datasets. The proposed method is compared with existing methods, such as domain adaptation

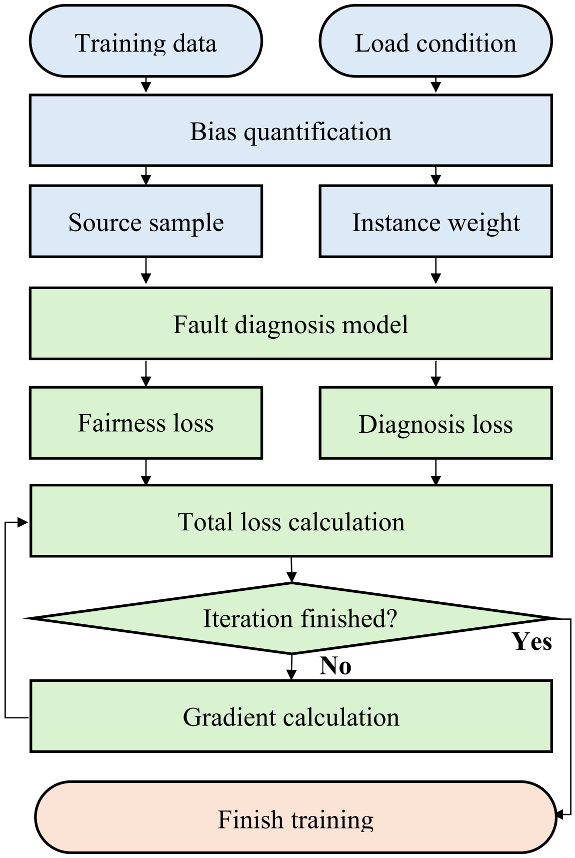


Figure 2. A flowchart of the proposed method

and fairness approaches. The performance is validated by quantitative analysis using accuracy and fairness criteria, and qualitative analysis using gradient-weighted class activation mapping (Grad-CAM).

4.1. Data Description

A public bearing dataset from the Society for Machinery Failure Prevention Technology (MFPT) is used for experimental validation (Bechhoefer, E., 2023). The detailed meta-information is described in Table 1. To evaluate the diagnosis performance under different load conditions, datasets A, B, and C are all used for training and test data. The number of training samples is 100, and the number of test samples is 100, considering insufficient training sample conditions in the general industrial field.

Table 1. A meta-information of the validation dataset. N is the normal state, IF is the inner race fault state, and OF is the outer race fault state.

Dataset	Speed	Load	Fault type
A	25 Hz	45.36 kgf	N/IF/OF
B	25 Hz	90.72 kgf	N/IF/OF
C	25 Hz	136.08 kgf	N/IF/OF

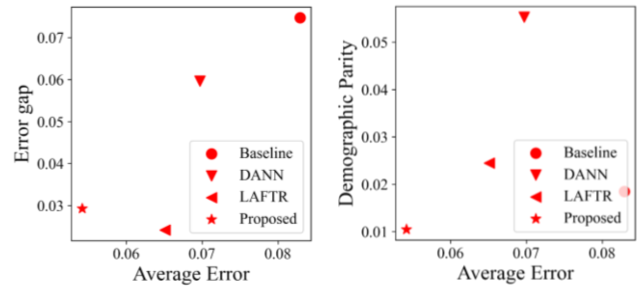


Figure 3. Comparative study for accuracy and diagnosis performances

4.2. Evaluation of Diagnosis and Fairness Performance

The diagnosis and fairness performance are evaluated simultaneously in Figure. 3. The proposed method can achieve superior diagnosis performance and fairness results. In Figure. 3, the x-axis is the average error, which is related to diagnosis performance, and the y-axis is the error gap and demographic parity, which means fairness-related criteria. The lower criteria mean superior performance in all cases. The proposed method shows similar fairness in error gap, and superior demographic parity compared with the Learning Adversarially Fair and Transferable Representation (LAFTR) technique (Madras, D., Creager, E., Pitassi, T., & Zemel, R., 2018). Moreover, the proposed method maintains a superior diagnosis performance compared to domain adversarial neural networks (DANN) and baseline 1D-CNN (Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., & Lempitsky, V., 2016). This implies that the proposed method is robust to bias under different load conditions due to the instance reweighting.

4.3. Bias Mitigation Performance

To validate the bias mitigation of the proposed method, Grad-CAM is performed for comparative models in Figure. 4. The existing fairness method, LAFTR, shows weak detection for low impulsive peaks in the signals. This is due to the impulsiveness bias from the high-impulsive instance. On the other hand, in Figure. 4(b), the proposed method highlights both weak and strong peaks in the signals. This means that the proposed method can mitigate the impulsiveness bias under different load conditions.

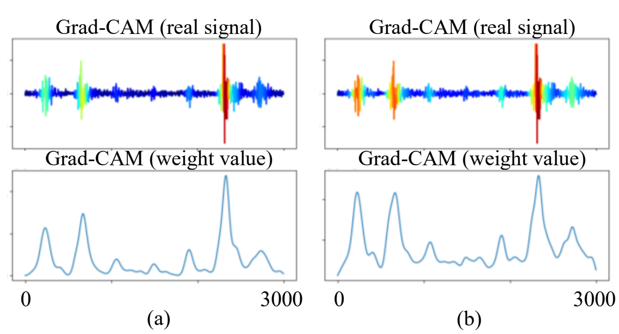


Figure. 4. Grad-CAM results for (a) existing fairness method, and (b) proposed method.

5. CONCLUSION

This research aims to mitigate the bias in the bearing diagnosis model under different load conditions due to impulsiveness. The proposed bias analysis can quantify the impulsive bias utilizing the normalized kurtosis values. These values are incorporated into a data-driven model using instance reweighting methods. The experimental results demonstrate that the proposed fairness-aware reweighting framework significantly improves both the accuracy and fairness of bearing fault diagnosis. The key to this success lies in reframing the problem. By treating the performance drop not as a domain shift, but as an internal model bias, the proposed method addresses the root cause of the error under different load conditions. The reweighting scheme compels the model to abandon its biased learning strategy of focusing only on salient, high-amplitude features and instead learn the fundamental physical patterns of a fault.

For future work, the proposed method can be more advanced by combining in-processing fairness algorithms or applying different fairness criteria suitable for specific diagnostic scenarios. For instance, in safety-critical applications, one might prioritize minimizing false negatives for severe faults, a goal that could be achieved by using a fairness criterion like Equalized Odds.

ACKNOWLEDGEMENT

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program- Automotive Industry Technology Development-Green Car) (RS-2024-00444961, Development and demonstration of Purpose-Built Electric vehicles using design platforms) funded by Ministry of Trade, Industry & Energy (MOTIE, Korea)

REFERENCES

Bechhoefer, E. Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms. *Available online:* https://www.mfpt.org/fault-data-sets/ (accessed on 30 October 2023).

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., &Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59), 1–35.

Kumar, K., Shukla, S., & Singh, S. K. (2022). Early detection of bearing faults using minimum entropy deconvolution adjusted and zero frequency filter. *Journal of Vibration and Control*, 28(9-10), 1011-1024.

Lee, J., Ko, J. U., Kim, T., Kim, Y. C., Jung, J. H., & Youn, B. D. (2024). Domain adaptation with label-aligned sampling (DALAS) for cross-domain fault diagnosis of rotating machinery under class imbalance. *Expert Systems with Applications*, 243, 122910.

Madras, D., Creager, E., Pitassi, T., & Zemel, R. (2018). Learning adversarially fair and transferable representations. *In International Conference on Machine Learning* (pp. 3384-3393). PMLR.

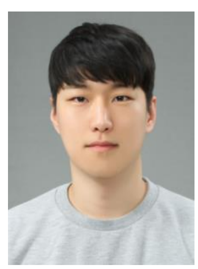
Ni, Q., Ji, J. C., Halkon, B., Feng, K., & Nandi, A. K. (2023). Physics-Informed Residual Network (PIResNet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing*, 200, 110544.

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—A tutorial.*Mechanical Systems and Signal Processing*, 25(2), 485–520.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In Artificial Neural Networks and Machine Learning–ICANN 2018: *27th International Conference on Artificial Neural Networks*, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27 (pp. 270-279). Springer International Publishing.

BIOGRAPHIES

Seungyun Lee is Ph.D. student of Mechanical Engineering at Seoul National University, Seoul, Republic of Korea. He received his B.S degree in biosystem engineering from Seoul National University in 2021. His research areas include artificial intelligence-based mechanical fault diagnosis, physics-informed artificial intelligence, and fault diagnosis of very low-speed bearings. His research efforts result in remarkable awards including Postech Open Innovation Big Data Challenge 2nd Place Winner (2021), Heart Disease AI Datathon Winner (2021), Industrial AI Hackathon Factory Hack Korea 2nd Place Winner (2023), and IEEE PHM Data Challenge Winner (2023).



Sungjong Kim is Ph.D. candidate student at the Seoul National University, Department of Mechanical Engineering in Laboratory for System Health and Risk Management. He received bachelor's degree in mechanical engineering from Seoul National University, Seoul, South Korea, in 2020. His recent research interest is deep learning based bearing fault diagnosis, explainable AI and low speed bearing fault diagnosis. He has been awarded Industrial AI Hackathon Factory Hack Korea (2021) 3rd place Winner and K-AI Manufacturing Data Challenge (2021) 3rd place Winner.

Minjae Kim is a Ph.D. student in Mechanical Engineering at Seoul National University, Seoul, Republic of Korea. He received his B.S. degree in Mechanical Engineering from Yonsei University in 2023. His recent research interests include domain-informed neural network-based bearing fault diagnosis and the development of artificial intelligence-based battery management systems. His research achievements include the Industrial AI Hackathon Factory Hack Korea 3rd Place Winner (2024).

Heonjun Yoon is an assistant professor in the School of Mechanical Engineering at Soongsil University, Seoul, Republic of Korea. Prof. Yoon was a visiting scholar in the George W. Woodruff School of Mechanical Engineering at the Georgia Institute of Technology, Atlanta, Georgia, USA, in 2019-2020. Prof. Yoon received a Ph.D. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Republic of Korea, in 2018. Prof. Yoon has explored three research avenues with great enthusiasm as: piezoelectric vibration energy harvesting, elastic metamaterial, and artificial intelligence-driven design.

Byeng D. Youn is a Professor of Mechanical Engineering at Seoul National University (SNU) and the founder and CEO of ONEPREDICT Corp. (onepredict.ai). He received his B.S. in Mechanical Engineering from Inha University, South Korea (1996), his M.S. from the Korea Advanced Institute of Science and Technology (KAIST, 1998), and his Ph.D. from the University of Iowa, USA (2001), all in Mechanical Engineering. His current research interests include industrial artificial intelligence, intelligent metrology, system modeling, and AI-based decision making. Professor Youn's contributions to research have earned wide peer recognition and numerous honors, including World's Top 2% Scholar, Stanford University (2020-2024), 0.05% of scholars in the design reliability and optimization by Scholar GPS, the Prime Minister Award (2019), the Outstanding Research and Service Awards from the Korean Society of Prognostics and Health Management (KSPHM), the Korean Society of Design Optimization (2019), the Shin Yang Academic Award from Seoul National University (2017), the 16-time winner of Global PHM Data Challenges including PHM Society Data Challenge. He currently serves as a Senior Editor for the journal Structural and Multidisciplinary Optimization.