4th Asia Pacific Conference of the Prognostics and Health Management,
Tokyo, Japan, September 11 – 14, 2023 OS05-06

Advancing Predictive Maintenance: A Study of Domain Adaptation for Fault Identification in Gearbox Components

Shinya Tsuruta[1], Koji Wakimoto[1], Takaaki Nakamura[1], Shahin Siahpour[2], Marcella Miller[2], John Taco[2], and Jay Lee[3]

[1] *Mitsubishi Electric corporation, Kamakura-shi, Kanagawa, 247-8501, Japan*
Tsuruta.Shinya@dw.MitsubishiElectric.co.jp
Wakimoto.Koji@df.MitsubishiElectric.co.jp
Nakamura.Takaaki@dy.MitsubishiElectric.co.jp

[2] *Department of Mechanical Engineering, University of Cincinnati, Cincinnati, OH, 45221, USA*
siahposn@mail.uc.edu
mille5mc@mail.uc.edu
tacolojo@mail.uc.edu

[3] *Department of Mechanical Engineering, University of Maryland, College Park, MD 20742, USA*
lj2@ucmail.uc.edu

Abstract

This paper explores the use of machine learning in predictive maintenance, which has been increasingly demanded in recent years to reduce downtime and maintenance burden. The challenge of different data distributions between training and test data in machine learning is common in predictive maintenance where equipment operation patterns can change, leading to reduced operational efficiency. The authors validate a domain-adaptive anomaly detection method combining CNN and MMD, which achieves similar accuracy with PCA, SVD, and other dimensionality reduction methods. The study also shows that the method maintains accuracy even when the number of normal data in the target domain is 1/10 of the source domain.

1. Introduction

Gearboxes are pivotal in industrial settings, critical for the functioning of wind turbines, conveyors, and industrial robots. Their health is vital, given their susceptibility to wear and corrosion, making their close monitoring essential to prevent unexpected downtime, production loss, and safety hazards. Early identification of gearbox health is therefore crucial.

Traditional gearbox prognostics and health management (PHM) typically use signal processing techniques applied to vibrational data(Qiao & Lu, 2015). These techniques, such as time synchronous average method(Bennett, 1958), gear mesh frequencies observation(Soualhi et al., 2018), and various domain approaches, have proven effective but require domain expertise and intensive feature extraction. Recently, neural networks, specifically convolutional ones(Jing, Zhao, Li, & Xu, 2017; Han, Liu, Yang, & Jiang, 2019; Grezmak, Wang, Sun, & Gao, 2019), have been applied to vibration signal analysis, with deep learning strategies also gaining traction(Jiang et al., 2017). These methods, able to perform automatic feature selection(Saufi, Ahmad, Leong, & Lim, 2019), are attractive due to their efficiency and the high information density of their extracted features.

However, deep learning's need for comprehensive data is a challenge, as obtaining data for every possible working condition is often unfeasible. To address this, transfer learning has been explored, allowing knowledge from one machine state to enhance the assessment of others. Yet, domain discrepancy presents an obstacle in transfer learning(Costa, Akçay, Zhang, & Kaymak, 2020), and domain adaptation (DA) methods must be used to maintain result accuracy(Siahpour, Li, & Lee, 2022).

In industrial settings, domain adaptation is typically performed on the healthy class, with deep learning methods then identifying anomalies that do not fall within the expected range. These outliers can signal necessary maintenance actions. To address aforementioned issues and challenges, this work studies the following novelties:

1. To mitigate the challenges of domain discrepancy in

transfer learning, we integrated domain adaptation methods within the dimensionality reduction process. This novel integration is aimed at enhancing the robustness and accuracy of fault identification across diverse operating conditions, even in the presence of domain shifts.

2. We pioneered a comprehensive comparative study on the effectiveness of different dimensionality reduction methods such as PCA, SVD, Isomap, and t-SNE, in the context of anomaly detection for gearbox health monitoring.
3. We investigated the influence of varying quantities of normal data in the target domain on the performance of our proposed approach, providing insights into optimal data usage for such applications.

The remaining sections of the paper are organized as follows: Section II reviews related literature, Section III presents the mathematical background, Section IV introduces the proposed methodology, Section V investigates the corresponding experimental study, and Section VI concludes.

2. Related works

Data-driven methods like deep learning, domain adaptation, and anomaly detection have gained popularity in intelligent systems due to advancements in AI. These techniques facilitate pattern recognition, dataset alignment, and classification, and their application has been highlighted in recent literature.

For instance, (Li, Ding, & Sun, 2018) used deep convolutional neural networks (DCNN) for aero-engine life prediction requiring no prior knowledge. (Zhou, Yang, Fujita, Chen, & Wen, 2020) tackled unbalanced data using a generative adversarial network (GAN), while (Liu et al., 2020) leveraged DCNN for gear grinding monitoring.

However, deep learning models' generalization can be compromised if training and test data differ, necessitating domain adaptation techniques. (B. Zhang, Li, Tong, & Zhang, 2017) used unsupervised domain adaptation for bearing fault detection, aligning training and test data distributions. (Buijs, Koch, & Dugundji, 2021) applied transfer learning to transportation mode choice with positive results, while (W. Zhang, Li, Ma, Luo, & Li, 2021) combined transfer learning with deep representation regularization for life prediction.

Anomaly detection is another effective fault diagnosis technique. (Purarjomandlangrudi, Ghapanchi, & Esmalifala, 2014) used data mining for bearing fault diagnosis, yielding a 95% accuracy. (Yang, Ma, Zeng, Peng, & Liu, 2021) improved spacecraft telemetry data anomaly detection using a long short-term memory (LSTM) architecture.

These techniques have been combined for improved results. (Mahyari & Locker, 2018) used transfer learning for robotic predictive maintenance, eliminating false alarms. (Vincent, Wannes, & Jesse, 2020) proposed a semi-supervised anomaly detection using relevant labels from a related anomaly detection task. Lastly, (Michau & Fink, 2021) tested unsupervised transfer learning (UTL) on various datasets, showcasing promising results.

3. Proposed method

3.1. Architecture Overview

Our proposed model (Fig.1) operates in two stages with four primary modules. Stage one involves a deep learning feature extractor that pulls high-level (HL) features from the frequency spectrum of collected data. These HL features are transferred to the domain adaptation module, which creates domain-invariant features. These features are trained on source data, encompassing all health classes (H-healthy, B-between, F-faulty), and target domain with only the healthy class. The goal is to derive features that are domain-invariant and indicative of the healthy class. A second stage is introduced to enhance the anomaly detection capacity of these features. Here, a one-class classifier uses the features for anomaly detection after a dimension reduction module simplifies the data, retaining necessary information (Fig.2).

3.2. Cross-domain Transfer Learning

The initial stage aims at creating healthy class-discriminative, domain-invariant features. To ensure healthy class discrimination, we use the cross-entropy loss function.

$$L_c = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{N_c} 1\{y_i = j\}\log(y'_{ij}) \tag{1}$$

where n, N_c, y_i, and y'_{ij} are the the number of samples, the number of classes in each domain, the label corresponding to the ith sample, and the predicted label corresponding to the ith sample and jth class, respectively.

To enable knowledge transfer from the source to the target domain, a domain adaptation strategy is employed. Given that the source and target data are gathered under varying regimes, affecting data distribution, we utilize the MMD metric to measure the discrepancy and address this issue. This MMD term is added to the total optimization goals as a loss function as follows:

$$L_{DA} = MMD^2(F_{source}, P_{target}) \tag{2}$$

Where the arguments F_{source} and F_{target} are the feature representations of the data corresponding to the source and target domains, respectively.

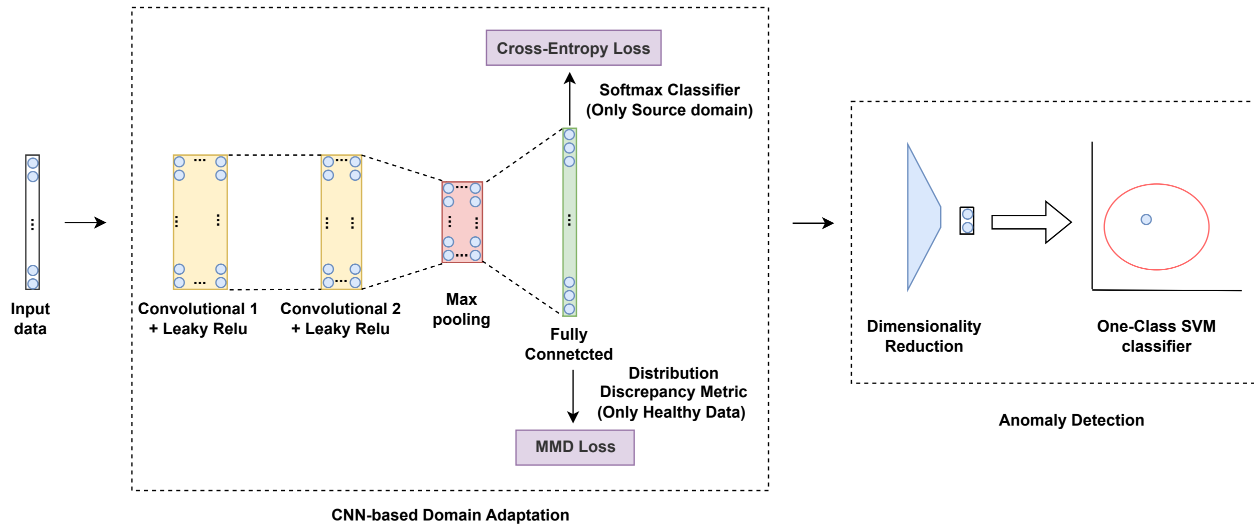

Figure 1. The schematic overview of the proposed two-stage anomaly detection architecture.

3.3. Network Optimization

The integration of these loss functions forms the network optimization goal. The general loss function is formed by integrating equations (1) and (2) as follows:

$$L_{tot} = \alpha L_c + \beta L_{DA} \tag{3}$$

Where $\alpha > 0$ and $\beta > 0$ are the penalty coefficients for L_c and L_{DA}, respectively.

During training, network parameters for the feature extractor and one-class classifier modules are optimized and updated based on the loss function as follows:

$$\theta \leftarrow \theta - \delta(\alpha \frac{\partial L_c}{\partial \theta} + \beta \frac{\partial L_{DA}}{\partial \theta}) \tag{4}$$

where δ indicates the learning rate.

3.4. High-level Anomaly Detection

The HL features collected from stage one are used for anomaly detection with the one-class classifier module. The features are mapped into 2D space using a dimensionality reduction algorithm before feeding to the module. Two hyper-parameters are tuned in this stage for the optimization of the RBF kernel hyper-parameter as follows:

$$k(\Phi_i, \Phi_j) = e^{-\gamma \|\Phi_i - \Phi_j\|_2} \tag{5}$$

where Φ_i, Φ_j represents the ith and jth feature samples and $\|.\|$ is the L2 norm operator.

Table 1. Detailed Data Collection Information.

Signal	Unit	Sampling rate
Command position	Encoder pulse count	444 μs
Feedback position	Encoder pulse count	444 μs
Command velocity	rpm	444 μs
Feedback velocity	rpm	444 μs
Feedback torque current	% of torque rate	444 μs
Total drive time	s	444 μs

Table 2. Specification of Network Parameters.

Parameter	Value	Parameter	Value
δ	1e-4	Epochs	500
Neurons of FC layer	128	Drop-out rate	0.5
Batch size for cross-entropy	32	Batch size fo MMD	100
Number of samples	1500		

4. EXPERIMENTAL STUDY

4.1. Data Profile

Our study analyzes gearbox data from an industrial system that's part of a planetary gearbox system (Fig.3). The studied gearbox and representative wear faulty condition of the gear are shown in Fig.4. The dataset includes both of healthy and faulty data. The faulty data was collected by grinding the gear tooth to reproduce war conditions.

The dataset consists of cycle features test (FCFT) data as outlined in (Liao & Lee, 2009), with our focus on steady state data over the transient portion. We collected 100 loops of 63 command patterns for each wear/gain condition. The first 48 patterns relate to four main working regimes of 50, 500, 1000, and 3000 rpm. For this study, we omitted some patterns and considered only forward rotation based on expert input, analyzing 100 loops of 32 patterns.

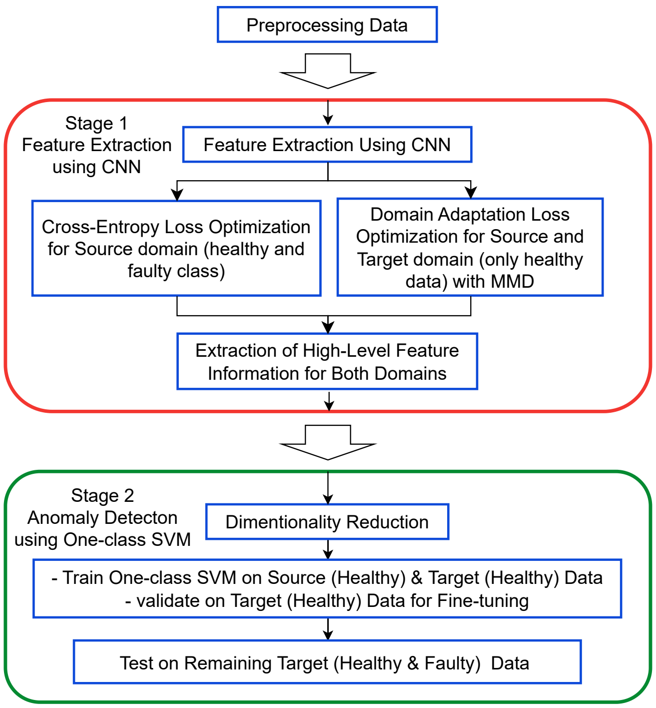

Figure 2. The flow chart explaining the proposed two-stage anomaly detection methodology.

Table 3. F1 score on each method.

Dimensionality reduction	F1 score
PCA	0.98
SVD	0.96
Isomap	0.96
t-SNE	0.811

Each loop comprises command and feedback position, velocity, and torque current, gathered every 444 μs . Supplementary data, like encoder temperature and load rate, is also collected. Table 1 summarizes the data collection information.

4.2. Data Preprocessing

Data processing begins by separating different loops and patterns, leading to a 100 loop x 32 pattern matrix. The aim is to truncate samples to contain only steady state data, achieved by retaining feedback torque current points corresponding to samples within one rpm of the working regime velocity. Further reduction removes overshoot, eliminating the first 40% of each sample to retain steady state data.

Post-cleanup, we concatenate the eight patterns from each loop and then concatenate these 100 loops. This results in a cleaned torque signal for each wear, gain condition, and working regime (36 signals in total).

The final step prepares the torque signals for our method. We downsample each signal, taking every 10th data point, and split it into 500 windows of 2000 time-domain points each. The windowing step size and overlap percentage are determined based on the total torque signal length.

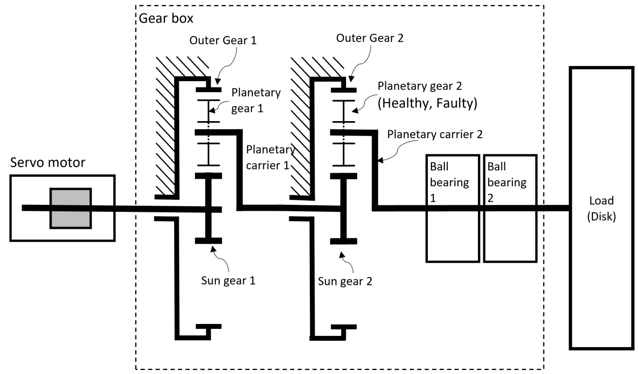

Figure 3. The experiment setup for the manufacturing system.

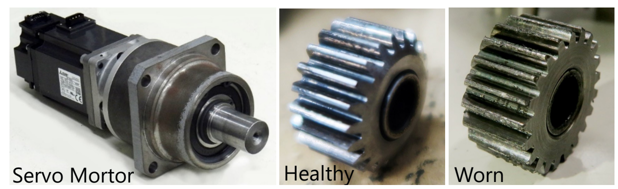

Figure 4. The studied gearbox and representative wear fault condition.

4.3. Implementation Details and Tasks

Several dimensionality reduction methods were used to evaluate the proposed method. PCA, SVD, Isomap, and t-SNE were used for dimensionality reduction. PCA and SVD are linear dimensionality reduction methods, while Isomap is a nonlinear dimensionality reduction method based on nonlinear distances on the manifold. t-SNE performs dimensionality reduction so that sets that are similar with high probability are in the neighborhood and sets that are different are in the distance. From the gearbox data set, 500 rpm and 1000 rpm data were extracted by preprocessing. The evaluation was performed with 500 rpm as the source domain and 1000 rpm as the target domain. The feature extractor module is comprised of two consecutive convolutional layers with 30 and 20 filters (filter size = 5), respectively. The leaky rectified linear unit (ReLU) is utilized as the activation function in both convolutional layers. The convolutional layers are followed by a max pooling layer with a pooling size of 2. The detailed information of the network implementation and parameters are provided in Table 2.

4.4. Comparison of Dimensionality Reductions

Table 3 shows the F1-score of anomaly detection for each dimensionality reduction method. There is no significant dif-

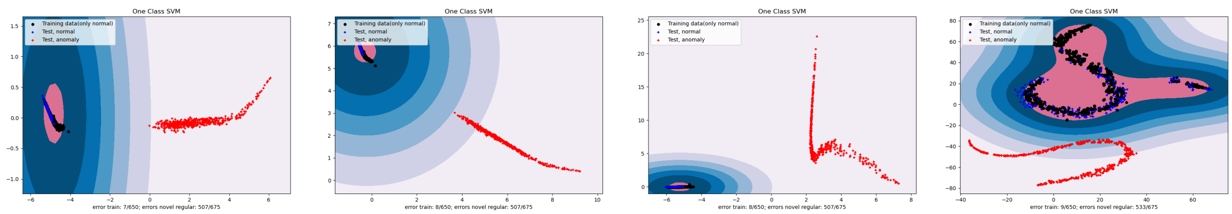

Figure 5. Comparison of Dimensionality Reductions.

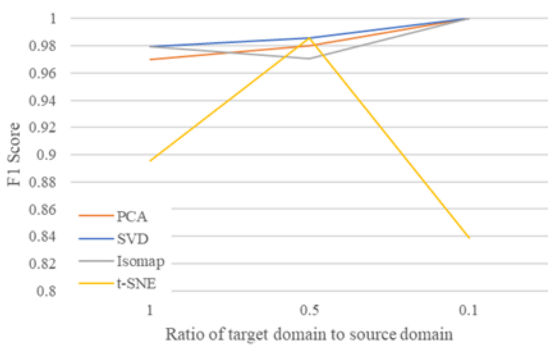

Figure 6. Reduction of normal data.

ference among each anomaly detection method, and they all show high accuracy. Compared to the other dimensionality reduction methods, t-sne has a lower score because the normal data in the source and target domains are not aggregated on the feature map, resulting in many false positives. Table3 shows the inference results for each dimensionality reduction method. That is, Fig.5 shows the two-dimensional feature map of the one-class SVM. As shown in the figure, it can be seen that normal and abnormal data can be separated. In the inference results using PCA, SVD, and Isomap, the normal and abnormal data are much further apart on the feature map compared to t-SNE.

4.5. Reduction of Normal Data

Fig.6 shows the accuracy of anomaly detection when the number of normal data in the target domain was reduced. t-SNE showed a higher number of false positives and lower accuracy. t-SNE showed almost no effect on accuracy when the number of normal data was reduced to 50 for PCA, SVD, and Isomap. The t-SNE is prone to over-fitting due to the nonlinearity in the feature map with dimensionality reduction, which is easily fitted to the nonlinear normal region by one-class SVM. On the other hand, the linear dimensionality reduction method is considered to be more robust because of the formation of nonlinear normal regions for linear feature distributions.

5. Conclusion

iIn this study, we compared dimensionality reduction methods in a two-stage deep learning-based transition learning method and experimentally verified their effectiveness in reducing normal data in the target domain. In the comparison of dimensionality reduction methods, the results were compared using t-SNE, and all dimensionality reduction methods showed high anomaly detection performance. In the target domain normal data reduction study, high accuracy was maintained in PCA, SVD, and Isomap even when the target domain normal data was reduced to 1/10 of the source domain normal data. Further validation is planned for application to industrial equipment with multiple rotation axes, such as robot manipulators.

Acknowledgment

This study was supported and funded by Mitsubishi Electric Corporation. We would like to express our gratitude to our colleague, Naoto Takano, Mitsubishi Electric Corporation, who provided us with gearbox data and many valuable insights about motion systems.

References

Bennett, W. (1958). Statistics of regenerative digital transmission. *Bell System Technical Journal*, *27*(6), 1501-1542.

Buijs, R., Koch, T., & Dugundji, E. (2021). Applying transfer learning and various ann architectures to predict transportation mode choice in amsterdam. *Procedia Computer Science*, *184*, 532-540.

Costa, P. R. d. O. d., Akçay, A., Zhang, Y., & Kaymak, U. (2020). Remaining useful lifetime prediction via deep domain adaptation. *Reliability Engineering & System Safety*, *195*, 106682.

Grezmak, J., Wang, P., Sun, C., & Gao, R. X. (2019). Explainable convolutional neural network for gearbox fault diagnosis. *Procedia CIRP*, *80*, 476-481.

Han, T., Liu, C., Yang, W., & Jiang, D. (2019). A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-based systems*, *165*, 474-487.

Jiang, G., Zhao, J., Jia, C., He, Q., Xie, P., & Meng, Z. (2017). Intelligent fault diagnosis of gearbox based on vibration and current signals: a multimodal deep learning approach. *Prognostics and System Health Management Conference (PHM-Qingdao)*, 1-6.

Jing, L., Zhao, M., Li, P., & Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, *11*, 1-10.

Li, X., Ding, Q., & Sun, J. Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, *172*, 1-11.

Liao, L., & Lee, J. (2009). A novel method for machine performance degradation assessment based on fixed cycle features test. *Journal of Sound and Vibration*, *326*(3-5), 894-908.

Liu, C., Mauricio, A., Chen, Z., Declercq, K., Meerten, Y., Vonderscher, Y., & Gryllias, K. (2020). Gear grinding monitoring based on deep convolutional neural networks. *IFAC-PapersOnLine*, *53*(2), 10324-10329.

Mahyari, A. G., & Locker, T. (2018). Domain adaptation for robot predictive maintenance systems. *arXiv preprint*, arXiv:1809.08626.

Michau, G., & Fink, O. (2021). Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems*, *216*, 106816.

Purarjomandlangrudi, A., Ghapanchi, A. H., & Esmalifala. (2014). A data mining approach for fault diagnosis: An application of anomaly detection algorithm. *Measurement*, *55*, 343-352.

Qiao, W., & Lu, D. (2015). A survey on wind turbine condition monitoring and fault diagnosis—part ii: Signals and signal processing methods. *IEEE Transactions on Industrial Electronics*, *62*(10), 6546-6557.

Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019). Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *Ieee Access*, *7*, 644-662.

Siahpour, S., Li, X., & Lee, J. (2022). A novel transfer learning approach in remaining useful life prediction for incomplete dataset. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1-11.

Soualhi, A., Hawwari, Y., Medjaher, K., Clerc, G., Hubert, R., & Guillet, F. (2018). Phm survey: implementation of signal processing methods for monitoring bearings and gearboxes. *International Journal of Prognostics and Health Management*, *9*(2), 0.

Vincent, V., Wannes, M., & Jesse, D. (2020). Transfer learning for anomaly detection through localized and unsupervised instance selection. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(4), 6054-6061.

Yang, L., Ma, Y., Zeng, F., Peng, X., & Liu, D. (2021). Improved deep learning based telemetry data anomaly detection to enhance spacecraft operation reliability. *Microelectronics Reliability*, 114311.

Zhang, B., Li, W., Tong, Z., & Zhang, M. (2017). Bearing fault diagnosis under varying working condition based on domain adaptation. *arXiv preprint*, arXiv:1707.09890.

Zhang, W., Li, X., Ma, H., Luo, Z., & Li, X. (2021). Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions. *Reliability Engineering & System Safety*, *211*, 107556.

Zhou, F., Yang, S., Fujita, H., Chen, D., & Wen, C. (2020). Deep learning fault diagnosis method based on global optimization gan for unbalanced data. *Knowledge-Based Systems*, *187*, 104837.

BIOGRAPHIES

Shinya Tsuruta received his M.S. degree in precision engineering from the University of Tokyo in 2011. From 2011 to 2019, he started his career as a camera hardware engineer and then contributed as an embedded software engineer to improve the performance of industrial printers and develop fault prediction systems. Since 2020, He has joined the R&D center at Mitsubishi Electric Corp., Japan, focusing on data analysis and machine learning. His current interests include anomaly detection, conditioned-based maintenance, and other applications of machine learning in manufacturing.

Koji Wakimoto received his B.S. and M.S. degrees from Kyoto University in 1983 and 1985 respectively and received his Dr. Eng. degree from the University of Tokyo in 1996. Since 1985 he has been with Mitsubishi Electric Corp., Japan. His research interests include data analysis, computer vision, and geospatial information systems. He is a member of the IEEE.

Takaaki Nakamura received his M.S degree from Kyushu University in 2002. Since 2002, he has been with Mitsubishi Electric Corp., Japan. His research interests include data analysis and machine learning.

Shahin Siahpour received his B.S. degree in mechanical engineering from Shiraz University, Shiraz, Iran, in 2015, and the M.S. degree in mechanical engineering from University of Tehran, Tehran, Iran, in 2017. He is currently pursuing the Ph.D. degree in mechanical engineering with the University of Cincinnati, Cincinnati, OH, USA. His research interests include deep learning, prognostics, health management, and industrial AI.

Marcella Miller received the B.S. and M.Eng. degrees in mechanical engineering from the University of Cincinnati, Yang, Cincinnati, OH, USA, in 2019, where she is currently pursuing the Ph.D. degree in mechanical engineering with the Department of Mechanical and Materials Engineering.

Her research interests include prognostics and health management, machine learning, and probabilistic methods.

John Taco received his B.S. degree in mechanical engineering from Pontificia Universidad Catolica del Peru, Lima, Peru, in 2018. He is currently pursuing the M.S degree in mechanical engineering with the University of Cincinnati, Cincinnati, OH, USA. His research interests include deep learning, prognostics, health management, and industrial AI.

Jay Lee received his D.Sc. degree from George Washington University, Washington, DC, USA. He is currently a Clark Distinguished Professor and Director of Industrial AI Center at Univ. of Maryland College Park. Previously, he served as Ohio Eminent Scholar, a L. W. Scott Alter Chair Professor, and a Distinguished University Professor with the University of Cincinnati, Cincinnati, OH, USA. He was the Founding Director of the National Science Foundation (NSF) Industry/University Cooperative Research Center on Intelligent Maintenance Systems, which is a multicampus NSF Industry/University Cooperative Research Center consisting of the University of Cincinnati (lead institution), the University of Michigan, Ann Arbor, MI, USA, the Missouri University of S&T, Rolla, MO, USA, and the University of Texas-Austin, Austin, TX, USA. Since its inception in 2001, the Center has been supported by more than 100 global companies and was ranked with the highest economic impact (270:1) by NSF Economics Impacts Report in 2012.