Multi-Class Gearbox Fault Diagnosis via Pre-Trained Model-based Domain Adaptation with Healthy-Only Target Data

Dai-Yan Ji[1], Hanqi Su[1], Shinya Tsuruta[2], Daichi Arimizu[2], Yuto Hachiya[2], Koji Wakimoto[2], and Jay Lee[1]

[1] *Center for Industrial Artificial Intelligence, Department of Mechanical Engineering,*
University of Maryland, College Park, MD, 20742, USA
jidn@umd.edu
hanqisu@umd.edu
leejay@umd.edu

[2] *Mitsubishi Electric Corporation, Kamakura-shi, Kanagawa, 247-8501, Japan*
Tsuruta.Shinya@dw.MitsubishiElectric.co.jp
Arimizu.Daichi@dp.MitsubishiElectric.co.jp
Hachiya.Yuto@ea.MitsubishiElectric.co.jp
Wakimoto.Koji@df.MitsubishiElectric.co.jp

Abstract

Accurate gearbox fault diagnosis under varying operational speeds is critical for industrial predictive maintenance. A significant challenge is domain shift, where models trained under one condition fail to generalize to another, especially when only healthy data from the target domain is available for training. This study proposes a novel domain adaptation framework, CDANet, that directly leverages raw sensor data to perform multi-class fault classification without manual feature engineering. The model combines a lightweight CNN-based temporal feature extractor with a frozen DistilBERT encoder to capture transferable, domain-invariant representations, combined with a maximum mean discrepancy loss to align the feature distributions between the source and target domains using only healthy samples. Experimental results demonstrate that our proposed model significantly outperforms conventional deep learning approaches, achieving high classification accuracy across six domain adaptation tasks. This work validates the effectiveness of applying pre-trained models in domain adaptation for gearbox fault diagnosis under real-world domain shift constraints.

1. Introduction

Prognostics and health management (PHM) has become a crucial part of maintenance strategies for industrial applications, significantly improving operational reliability, safety, and cost-effectiveness. Specifically, effective fault diagnosis in gearbox systems is essential due to their significant impact on system performance. Failures in gearboxes can result in substantial downtime and increased maintenance costs. Recent reviews indicate that PHM helps reduce unexpected breakdowns, optimize maintenance schedules, and accurately predict component lifespan, thus providing tangible economic and operational benefits (Soualhi et al., 2018; Huang et al., 2024).

Traditionally, machine learning (ML) techniques in gearbox fault diagnosis heavily rely on feature extraction methodologies to capture essential diagnostic information from collected data. Time-domain features, such as root mean square (RMS), kurtosis, skewness, and peak values, have been widely adopted and proven particularly effective in identifying various gearbox faults. Extensive extraction and selection of these features can significantly impact diagnostic performance, particularly in scenarios with limited data availability (Kumar, Singh, Kumar, & Sarangi, 2025). Moreover, additional systematic feature design approaches have further demonstrated their effectiveness in improving calibration and diagnostic accuracy across various machinery applications, including semiconductor manufacturing (Ji, Sumiya, et al., 2025).

Recent advancements in deep learning (DL) have further enhanced diagnostic capabilities, particularly in planetary gearbox systems. DL methods offer automatic extraction of fault features directly from raw vibration signals, reducing the dependence on domain-specific expertise and facilitating more reliable and generalized fault diagnosis. A notable approach integrates convolutional neural networks (CNNs) with transformer models, leveraging both local feature extraction and global sequence modeling, resulting in robust fault diagnosis capabilities that achieve high accuracy and reliability (Ji, Minami, & Lee, 2025).

However, a significant challenge that persists in real-world scenarios is domain shift, where operational conditions and system behaviors differ substantially between training and testing environments. Domain shift can severely degrade model performance, prompting extensive research into domain adaptation (DA) methods. DA is particularly valuable in addressing discrepancies in operational behaviors and conditions between the source and target domains, improving the robustness and generalization of diagnostic models. Recent studies have proposed various DA methodologies, such as adversarial learning, adaptive domain adversarial networks, conditional domain adversarial networks, and maximum mean discrepancy-based approaches (Ahmad et al., 2024).

In addition to these established approaches, this study explores the potential of leveraging pre-trained model-based domain adaptation techniques, particularly in scenarios where only healthy samples from the target domain are available, as a promising direction for enhancing multi-class gearbox fault diagnosis. While large language models (LLMs), such as DistilBERT (Sanh, Debut, Chaumond, & Wolf, 2019; Shin, Park, Baek, & Kim, 2023), are originally designed for textual data, this study explores a novel application of their internal components—specifically, the feed-forward sublayers, residual connections, and layer normalization operations trained during pretraining. By adapting these transferable components to non-textual time series data, the proposed approach demonstrates a new direction in cross-domain representation learning. This enables the model to generalize effectively across differing operational conditions, even with limited labeled information from the target domain. To achieve this, the model is trained using a joint loss function comprising two components: (1) a multi-class cross-entropy loss computed using all labeled samples from the source domain, and (2) a domain adaptation loss based on Maximum Mean Discrepancy (MMD), calculated exclusively from healthy class samples in both source and target domains. The MMD loss encourages the alignment of feature distributions between domains and mitigates the impact of domain shift. The research addresses a partial domain adaptation scenario and contributes toward more robust and deployable diagnostic frameworks in real-world gearbox health monitoring applications.

The remainder of this paper is shown as follows. Section 2 describes an overview of related work. Section 3 details the proposed methodology. Section 4 illustrates the experimental tasks and evaluates the performance. Finally, Section 5 concludes the paper.

2. RELATED WORKS

DA has emerged as a crucial solution for addressing the challenges posed by domain shift, significantly impacting machinery diagnostics and prognostics. In industrial machinery applications, failures due to faults in critical components, such as bearings and gearboxes, greatly influence operational safety and reliability. However, machine learning-based diagnostic models often struggle with limited labeled data and poor generalization across varied operational conditions, necessitating robust DA methods to enhance their applicability and accuracy (Yao, Kang, Zhou, Rawa, & Abusorrah, 2023).

Several recent domain adaptation methodologies have been developed to address these limitations. (Ye, Yan, Jiang, & Chen, 2025) introduced a multi-branch attention coupled convolutional domain adaptation network (MACCDAN) designed for intelligent fault recognition under unlabeled sample scenarios. MACCDAN employs a cross-attention coupled module and global feature aggregation, complemented by a maximum-similarity minimum-discrepancy adversarial loss to enhance cross-domain feature alignment. Similarly, (Shao, Jiang, Zhang, Zhou, & Huang, 2025) proposed a pseudo-label progressive learning guided wavelet class-aware adaptive network, leveraging discrete wavelet transforms and pseudo-label refinement to improve cross-domain fault diagnosis performance for gearboxes. Additionally, (Khanal et al., 2025) developed a domain-specific dual network utilizing multisource domain data, combining deep convolutional neural networks with convolutional block attention modules for improved transfer fault prognosis.

Domain adaptation techniques have also been integrated into digital twin frameworks to further address data scarcity issues. (Zhu, Deng, Tang, Yang, & Li, 2025) proposed a digital twin-enabled entropy regularized wavelet attention domain adaptation network specifically for gearbox fault diagnosis, which employs entropy regularization and wavelet-based feature extraction to bridge discrepancies between simulation and real operational data. Similarly, (Zhang, Li, & Wang, 2025) introduced a joint domain-adaptive transformer for remaining useful life (RUL) prediction, effectively aligning global feature distributions across different domains.

Moreover, the challenge of data scarcity and the difficulty of acquiring real-world fault data have led researchers to explore simulation-to-real (S2R) domain adaptation approaches. (Ji, Wang, Inoue, & Kanemaru, 2025) developed a hybrid physics-based and data-driven framework, employing topological data analysis and domain adaptation techniques to bridge gaps between simulated and real motor fault data. (Wang, Taal, & Fink, 2021) proposed integrating expert knowledge with domain adaptation, utilizing synthetic fault data generated through expert-informed methods, subsequently adapted to real data using an imbalance-robust DA approach. (Lou, Kumar, & Xiang, 2022) similarly employed finite element method (FEM)-based simulation data combined with generative adversarial network (GAN)-based domain adaptation, effectively reducing discrepancies between simulated and measured fault signals for machinery diagnosis. These advancements underscore the critical role of DA methods in machinery diagnostics, addressing domain shift challenges and enhancing diagnostic model reliability and robustness in real-world industrial applications.

3. METHODOLOGY

3.1. Problem Statement

Gearbox fault diagnosis is essential in maintaining operational reliability and safety of machinery systems. The primary components of interest for this diagnosis include gears and bearings. Fault conditions, such as gear wear and bearing corrosion, significantly degrade system performance and increase maintenance costs. Specifically, gear wear typically results from prolonged mechanical friction and contact stress within the gearbox, while bearing corrosion is often attributed to environmental factors such as moisture and contaminants.

Formally, we define our source domain dataset as $D_s = \{(x_s^i, y_s^i)\}_{i=1}^{n_s}$, consisting of labeled samples across three classes: healthy, gear wear, and bearing corrosion. The target domain dataset is represented as $D_t = \{(x_t^j, y_t^j)\}_{j=1}^{n_t}$. Importantly, during training, we only utilize a small subset of the target domain data, D_t^{train}, which contains exclusively healthy class samples. The remainder of the unseen target domain data, D_t^{test}, encompassing all three classes, is used to evaluate the model's multi-class classification performance.

3.2. Data Collection and Preprocessing

An industrial planetary gearbox system subjected to different operational conditions was used to collect this dataset. The experimental setup is shown in Figure 1. The dataset comprises three classes: (1) healthy samples, (2) faulty samples due to gear wear at a planetary gear in the gearbox, and (3) faulty samples due to corrosion at a ball bearing in the servo motor.

The data used in this study are collected under three working regimes corresponding to rotational speeds of 500, 1000, and 3000 RPM. For each speed regime, 12 distinct command patterns under high servo gain settings are executed, totaling 36 patterns across all conditions. Each command pattern, including both forward (positive) and backward (negative) rotations, is repeated 100 times for the healthy and gear wear classes, and 88 times for the bearing corrosion class. The recorded signals comprise feedback torque current, command velocity, feedback velocity, command position, and feedback position, all sampled at a fixed interval of 444 µs. In this study, the feedback torque current is utilized as the primary signal for analysis. The raw feedback torque current contains

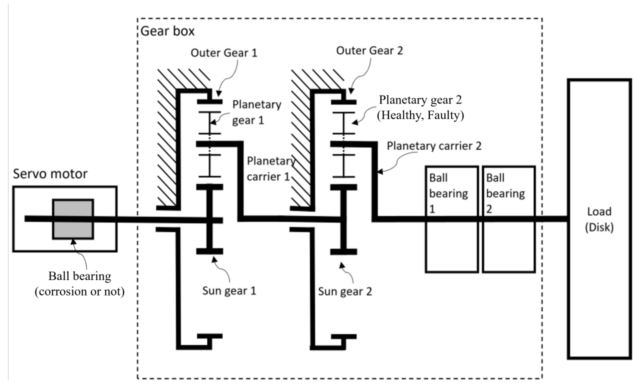

Figure 1. The experimental configuration for the manufacturing system.

both transient and steady-state components. To capture stable machine behavior, only the steady-state portion of the feedback torque current signal is extracted and truncated to a fixed length of 4000 sampling points.

3.3. Framework Architecture

In this study, the proposed CNN-DistilBERT Network (CDANet) model, is to perform time series classification by combining a convolutional neural network (CNN) with a pretrained DistilBERT transformer encoder. The overall architecture is illustrated in Figure 2.

The model takes a univariate time series as input, represented as a one-dimensional vector of length . A lightweight CNN-based feature extractor is employed to learn local temporal patterns. It consists of two 1D convolutional layers. The first convolutional layer uses 16 filters with a kernel size of 3 and padding of 1, followed by a ReLU activation. The second convolutional layer uses 32 filters with the same configuration and is also followed by a ReLU activation. The output feature maps are average-pooled along the temporal dimension to produce a fixed-length 32-dimensional feature vector. This vector is then projected to a 768-dimensional embedding space using a fully connected layer, aligning with the hidden dimension of the DistilBERT model. The projected embedding is reshaped to simulate a single-token sequence and is directly passed into the pretrained DistilBERT encoder. Due to the single-token input, the self-attention mechanism in the transformer layers is bypassed, and the model primarily utilizes the feed-forward sublayers, along with residual connections and layer normalization operations, within each transformer block. The DistilBERT parameters are frozen during training to reduce computational cost and prevent overfitting. The output of the encoder corresponds to the first token representation (analogous to a [CLS] token in BERT-based architectures), which is a 768-dimensional vector is extracted and passed through a final fully connected classification layer to generate class logits.

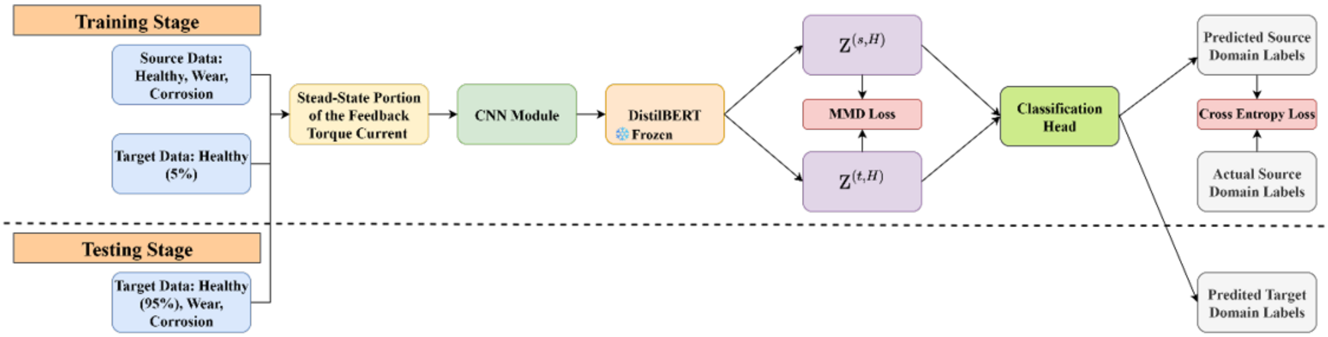

Figure 2. Architecture of the proposed CNN-DistilBERT Adaptation Network (CDANet) model.

To address domain shift issues in fault diagnosis, our model employs a joint optimization strategy integrating classification and domain discrepancy losses. Specifically, the classification loss, denoted as L_c, uses labeled samples from source and target domains, formulated as:

$$L_c = \frac{1}{N_s} \sum_{i=1}^{N_s} \sum_{c=1}^{C} 1\{y_i = c\} \log(\hat{y}_{i,c}) \tag{1}$$

Here, N_s is the total number of samples, C represents the number of classes, y_i is the actual class of the i^{th} sample, $\hat{y}_{i,c}$ denotes the predicted probability of class c for the i^{th} sample, and $1\{\cdot\}$ is the indicator function.

To mitigate domain discrepancy, particularly for healthy operational states, the Maximum Mean Discrepancy (MMD) measure is applied. Let $Z^{(s,H)}$ and $Z^{(t,H)}$ denote feature vectors extracted from the DistilBERT's final layer for healthy samples in source and target domains. The MMD loss L_{MMD} is given by:

$$L_{MMD} = MMD^2(Z^{(s,H)}, Z^{(t,H)}) \tag{2}$$

where $MMD^2(\cdot,\cdot)$ quantifies feature distribution discrepancies within a reproducing kernel Hilbert space (RKHS).

Finally, the overall loss function integrates these two losses as:

$$L_{total} = \alpha L_c + \beta L_{MMD} \tag{3}$$

Hyperparameters $\alpha > 0$ and $\beta > 0$ control the balance between achieving accurate classification and maintaining domain invariance, ensuring robust fault diagnosis across diverse operational conditions.

During training stage, the network parameters can be trained in each epoch as:

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

where θ represents the parameters of the network model, and δ is the learning rate.

Table 1. Data Distribution in Training and Testing Stages.

Stage	Data	Gearbox		Bearing
		Healthy	**Wear**	**Corrosion**
Training	Source Data	1200	1200	1056
	Target Data for DA	60	0	0
Testing	Target Data	1140	1200	1056

4. RESULT AND DISCUSSION

4.1. Experimental tasks

Each of the source and target domains comprises 3456 samples shown in Table 1, consisting of 1200 healthy gearbox samples, 1200 faulty gearbox samples (exhibiting wear), and 1056 faulty bearing samples (exhibiting corrosion). For domain adaptation, 5% (60 samples) of the healthy samples from the target domain are integrated into the source domain during the training phase. The remaining target domain samples are reserved exclusively for the testing phase. The training set is partitioned, with 90% of the data allocated for model training and the remaining 10% for validation. A fixed random seed is used during sampling to ensure reproducibility. To evaluate the model's performance under domain adaptation, we designed six tasks based on differences in rotational speed between the source and target domains. These rpm differences are set at 500, 1000, and 3000. A summary of these tasks can be found in Table 2. The hyperparameter settings for the proposed CDANet model are detailed as follows. The model is trained using a batch size of 8 for 100 epochs. The initial learning rate is set to 0.0008 and optimized using the Adam optimizer. A learning rate scheduler is employed with a step decay strategy, where the learning rate is multiplied by a factor of 0.96 every 5 epochs. The loss function in CDANet consists of two components: the classification loss and the domain adaptation loss. The weighting coefficients are set as $\alpha = 1.0$ for the cross-entropy loss and $\beta = 0.1$ for the MMD loss.

Table 2. Overview of the Designed Experiments.

Domain Adaptation Task	Source Domain	Target Domain
T1	500 rpm	1000 rpm
T2	500 rpm	3000 rpm
T3	1000 rpm	500 rpm
T4	1000 rpm	3000 rpm
T5	3000 rpm	500 rpm
T6	3000 rpm	1000 rpm

Table 3. Accuracy (%) on Each Domain Adaptation Task Across Different Models.

Task	Model (Accuracy %)		
	CNN	CNN+MMD	CNN+MMD+DistilBERT
T1	71.17	75.80	92.46
T2	70.97	74.85	92.29
T3	74.76	76.03	91.17
T4	96.35	96.17	98.88
T5	85.66	79.59	96.05
T6	96.79	99.20	99.44

4.2. Performance Evaluation

From Table 3, we first observe that the baseline CNN model DA exhibits relatively high accuracy on tasks T4 and T6, achieving 96.35% and 96.79%, respectively. Both tasks involve DA between 1000 rpm and 3000 rpm, suggesting that the signal characteristics between these two operating conditions share inherent similarities that the model can leverage effectively, even without explicit domain alignment.

When incorporating MMD into the CNN architecture (CNN+MMD), performance shows slight improvement on most tasks. For example, the accuracy increases from 71.17% to 75.80% in T1, from 70.97% to 74.85% in T2, and from 74.76% to 76.03% in T3. These results show that MMD contributes to mitigating domain shift by partially aligning the feature distributions between source and target domains. However, in T5, the performance decreases from 85.66% (CNN) to 79.59% (CNN+MMD), indicating that MMD-based DA may not be effective under certain challenging transfer scenarios.

When incorporating MMD into the CNN architecture (CNN+MMD), performance shows slight improvement on most tasks. For example, the accuracy increases from 71.17% to 75.80% in T1, from 70.97% to 74.85% in T2, and from 74.76% to 76.03% in T3. However, in T5, the performance decreases from 85.66% (CNN) to 79.59% (CNN+MMD). These results show that CNN+MMD contributes to mitigating domain shift by partially aligning the feature distributions, as evidenced in T1, T2, T3, and T6. In contrast, under certain challenging transfer scenarios, such as T4 and T5, this model may not be effective in improving performance.

The proposed CDANet (CNN+MMD+DistilBERT) model achieves the highest accuracy across all DA tasks. In particular, tasks with large domain gaps—such as T1, T2, and T3—benefit substantially from this architecture. CDANet improves the accuracy from 71.17% (CNN) and 75.80% (CNN+MMD) to 92.46% in T1, from 70.97% and 74.85% to 92.29% in T2, and from 74.76% and 76.03% to 91.17% in T3. Even for T4 and T6, where CNN already performs well, CDANet further boosts performance to 98.88% and 99.44%, respectively. Most notably, in task T5—where the performance of CNN+MMD drops below the baseline—CDANet significantly enhances the accuracy to 96.05%, highlighting its superior ability to adapt across complex and imbalanced domain shifts.

To better understand the performance of the proposed CDANet model, the confusion matrices for tasks T1–T6 are analyzed in Figure 3. The proposed CDANet model performs well overall, particularly in tasks T4 and T6, where all classes are classified with minimal errors. However, in tasks such as T1 and T2, a noticeable number of *Healthy* samples are misclassified as *Wear* and *Corrosion* classes, and a substantial portion of *Corrosion* samples are misclassified as *Wear*. Task T3 shows the reverse trend, with many *Wear* samples (239 out of 1200) misclassified as *Corrosion*. Task T5 also exhibits moderate confusion, particularly with *Wear* samples being misclassified as either *Healthy* (22 samples) or *Corrosion* (84 samples), and 28 *Corrosion* samples incorrectly predicted as *Healthy*.

5. CONCLUSION

This study proposes a novel framework, the CNN-DistilBERT Adaptation Network (CDANet) model, that integrates CNN feature extraction, a frozen DistilBERT encoder, and MMD-based alignment to address multi-class gearbox fault diagnosis under real-world domain shift conditions. The proposed method is particularly effective when the target domain contains only healthy samples during training—a common challenge in industrial scenarios. Through comprehensive experiments across six domain adaptation tasks, CDANet consistently outperforms conventional CNN and CNN+MMD models, achieving substantial improvements in classification accuracy, especially in tasks involving large domain gaps. These findings underscore the value of pre-trained architectures for time-series diagnostics and reinforce the effectiveness of integrating domain alignment and robust representation learning to address partial domain adaptation challenges in real-world industrial scenarios.

ACKNOWLEDGMENT

We would like to express our gratitude to our colleague, Naoto Takano, Mitsubishi Electric Corporation, who pro-

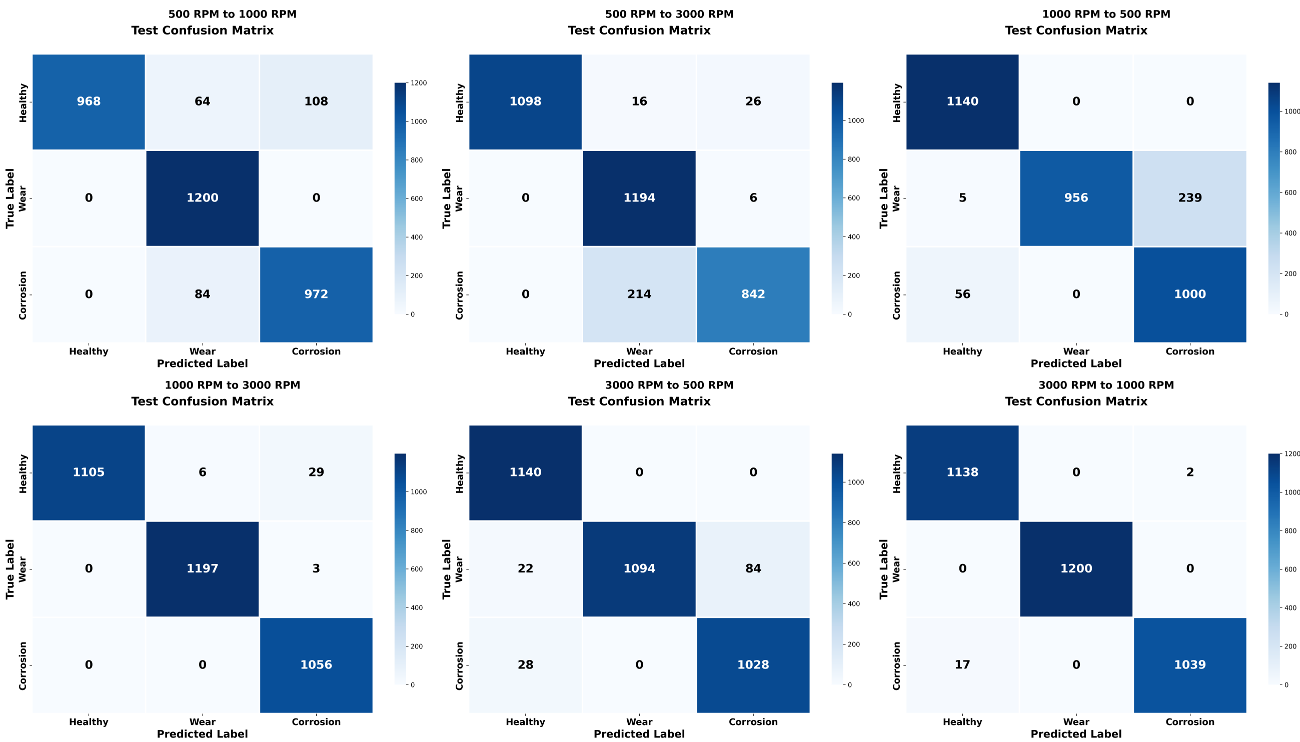

Figure 3. Confusion matrices of the proposed CDANet model on test data for domain adaptation tasks T1–T6.

vided us with gearbox data and many valuable insights about motion system.

REFERENCES

Ahmad, H., Cheng, W., Xing, J., Wang, W., Du, S., Li, L., … Lu, J. (2024). Deep learning-based fault diagnosis of planetary gearbox: A systematic review. *Journal of Manufacturing Systems*, *77*, 730–745.

Huang, C., Bu, S., Lee, H. H., Chan, C. H., Kong, S. W., & Yung, W. K. (2024). Prognostics and health management for predictive maintenance: A review. *Journal of Manufacturing Systems*, *75*, 78–101.

Ji, D.-Y., Minami, T., & Lee, J. (2025). Intelligent fault diagnosis for planetary gearboxes using a hybrid cnn-transformer model with skip connections. In *2025 ieee international conference on prognostics and health management (icphm)* (p. 1-6). doi: 10.1109/ICPHM65385.2025.11062027

Ji, D.-Y., Sumiya, M., Kamaji, Y., Matsukura, S., Li, W., & Lee, J. (2025). Improving machine calibration performance through systematic feature design in semiconductor manufacturing. In *2025 36th annual semi advanced semiconductor manufacturing conference (asmc)* (pp. 1–6).

Ji, D.-Y., Wang, B., Inoue, H., & Kanemaru, M. (2025). Motor fault detection with a hybrid physics-based and data-driven method. In *2025 ieee international electric machines drives conference (iemdc)* (p. 1178-1183). doi: 10.1109/IEMDC60492.2025.11061000

Khanal, S., Huang, H.-Z., Huang, C.-G., Dahal, A., Huang, T., & Niazi, S. G. (2025). Domain-specific dual network with unsupervised domain adaptation for transfer fault prognosis across machines using multiple source domains. *IEEE Transactions on Instrumentation and Measurement*.

Kumar, S., Singh, O. P., Kumar, V., & Sarangi, S. (2025). Ml-based bevel gearbox fault diagnosis: An extensive time domain feature extraction approach with limited data. *International Journal of Non-Linear Mechanics*, *170*, 105003.

Lou, Y., Kumar, A., & Xiang, J. (2022). Machinery fault diagnosis based on domain adaptation to bridge the gap between simulation and measured signals. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–9.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Shao, Z., Jiang, H., Zhang, X., Zhou, J., & Huang, W. (2025). Pll-wcan: Pseudo-label progressive learning guided wavelet class-aware adaptive network for gearbox cross-domain fault diagnosis. *Mechanical Systems and Signal Processing*, *230*, 112624.

Shin, C.-Y., Park, J.-T., Baek, U.-J., & Kim, M.-S. (2023). A feasible and explainable network traffic classifier utilizing distilbert. *IEEE Access*, *11*, 70216–70237.

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.

Wang, Q., Taal, C., & Fink, O. (2021). Integrating expert knowledge with domain adaptation for unsupervised fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–12.

Yao, S., Kang, Q., Zhou, M., Rawa, M. J., & Abusorrah, A. (2023). A survey of transfer learning for machinery diagnostics and prognostics. *Artificial Intelligence Review*, *56*(4), 2871–2922.

Ye, M., Yan, X., Jiang, D., & Chen, N. (2025). A multi-branch attention coupled convolutional domain adaptation network for bearing intelligent fault recognition under unlabeled sample scenarios. *Applied Soft Computing*, *174*, 113053.

Zhang, X., Li, Z., & Wang, J. (2025). Joint domain-adaptive transformer model for bearing remaining useful life prediction across different domains. *Engineering Applications of Artificial Intelligence*, *159*, 111615.

Zhu, P., Deng, L., Tang, B., Yang, Q., & Li, Q. (2025). Digital twin-enabled entropy regularized wavelet attention domain adaptation network for gearboxes fault diagnosis without fault data. *Advanced Engineering Informatics*, *64*, 103055.

BIOGRAPHIES

Dai-Yan Ji received his B.S. degree in Electronic Engineering and M.S. degree in Communications Engineering from Feng Chia University, Taiwan, in 2009 and 2012, respectively. He is currently pursuing a Ph.D. degree in Mechanical Engineering at the University of Maryland, College Park, MD, USA. His research interests include machine learning, signal processing, industrial artificial intelligence, and prognostics and health management.

Hanqi Su received his B.Eng. degree in Robotics Engineering from Southern University of Science and Technology, Shenzhen, Guangdong, China, in 2023. Previously, he participated in a one-year Special Student Program at the Massachusetts Institute of Technology (MIT) during his fourth undergraduate year. He is currently pursuing his Ph.D. in Mechanical Engineering at the University of Maryland, College Park. His research is primarily focused on machine learning, deep learning, prognostics and health management, engineering design, and industrial artificial intelligence.

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 system. 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 the other application of machine learning in manufacturing.

Daichi Arimizu received his B.S. and M.S. degrees from Kyoto University in 2020 and 2022 respectively. His master's research was in hadron physics and he conducted the development of semiconductor-based electron and positron detectors. Since 2022, he has been with Mitsubishi Electric Corp., Japan. His current research interests include data analysis and physics-informed AI.

Yuto Hachiya received his B.S. degree from Tokyo University of Science in 2019 and his M.S. degree from Tohoku University in 2021. Since 2021, he has been with Mitsubishi Electric Corp., Japan, focusing on the digitalization of production sites from 2021 to 2023. Since 2024, he has been engaged in the development of technologies for detecting anomalies in machinery. His research interests include data analysis, anomaly detection, and the estimation of remaining useful life.

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.

Jay Lee received his B.S. degree in mechanical engineering from Tamkang University, Taipei City, Taiwan, in 1979, M.S. degree in mechanical engineering from the University of Wisconsin-Madison, Madison, WI, USA in 1983, M.S. degree in industrial management from the State University of New York-Stony Brook, Stony Brook, NY, USA in 1987, and D.Sc. degree in mechanical engineering from George Washington University, Washington, DC, USA in 1992. Currently, he is a Clark Distinguished Chair Professor in Mechanical Engineering at the University of Maryland, College Park, and the founding director of the Industrial Artificial Intelligence Center to tackle the opportunities and unmet needs in the application of Artificial Intelligence to a wide range of industrial applications. Previously, he served as the Founding Director of the NSF I/UCRC on Intelligent Maintenance Systems (IMS) which has developed research memberships with over 100 global companies since 2000 and was selected as the most economically impactful I/UCRC in the NSF Economic Impact Study Report in 2012. He advised his students to win 1st place 5 times in the PHM Data Challenge during 2008-2016 as well as 1st Place in the PHM AP Data Challenge in 2023.