# Physics-Guided Sparse Domain Adaptation for Robust Rotor Defect Detection: Integrating System Identification, Synthetic Data Generation, and Maximum Mean Discrepancy Transfer Learning

Anil Kumar[1,*], Jianlong Wang[1], Rajesh Kumar[2]

[1] College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China
[2] Department of Mechanical Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India
* Corresponding author: anil.kumar@wzu.edu.cn

## Abstract

Rotor machinery underpins critical industrial processes across power generation, petrochemical, and aerospace sectors, yet experimental investigation of fault conditions is frequently impractical due to operational safety constraints, prohibitive costs of controlled fault induction, and the rarity of specific fault types in operational data. This paper proposes a physics-guided sparse domain adaptation (PGDL-SDA) framework that resolves the fundamental data scarcity challenge in rotor defect detection by integrating three complementary components. First, a rotor dynamic model is developed through system identification that extracts modal parameters (eccentricity and bearing damping coefficients) from limited experimental observations, enabling physics-faithful synthetic signal generation across a broad range of defect severities and operating conditions. Second, a hybrid deep learning training strategy combines labeled synthetic data from physics-based dynamic simulations with unlabeled experimental data from a physical test rig, exploiting the large labeled synthetic dataset while adapting toward the real-world signal distribution. Third, sparse domain adaptation employs Maximum Mean Discrepancy (MMD) minimization to reduce the distributional shift between synthetic and experimental feature spaces, Domain-Adversarial Neural Networks (DANN) for domain-invariant feature extraction, and entropy regularization to enforce sparse, confident predictions on unlabeled target-domain samples. Experimental evaluation on a custom rotor test rig with four fault conditions (normal, mass unbalance, shaft misalignment, and bearing outer race fault) demonstrates classification accuracy of 94.8% at 1500 RPM, outperforming CNN (81.2%), LSTM (82.1%), CNN+MMD (84.7%), DANN (86.3%), and non-sparse PGDL (89.5%) baselines. Cross-speed generalization experiments confirm the framework maintains above 91% accuracy under speed variations of +/-25%, establishing PGDL-SDA as a robust and practical solution for industrial rotor condition monitoring without requirement for large experimental fault datasets.

Keywords: physics-guided deep learning; rotor defect detection; system identification; domain adaptation; maximum mean discrepancy; DANN; synthetic data generation; condition monitoring

## 1. Introduction

Rotating machinery represents the backbone of energy conversion and process industries, with rotor systems constituting critical components in turbines, compressors, pumps, and electric motors that collectively account for a significant fraction of industrial capital assets worldwide [1,2]. Unscheduled rotor failures caused by mass

unbalance, shaft misalignment, bearing defects, and rotor crack propagation impose enormous costs through production downtime, emergency maintenance, and secondary equipment damage [3,4]. Condition-based maintenance (CBM) programs that continuously monitor rotor vibration signals and identify fault conditions before critical failure can dramatically reduce these costs while improving operational safety [5,6].

Data-driven fault diagnosis methods based on machine learning and deep learning have demonstrated impressive diagnostic performance on benchmark datasets, with convolutional neural networks (CNNs), recurrent architectures (LSTM, GRU), and attention mechanisms achieving classification accuracies exceeding 98% on public bearing fault datasets such as CWRU, MFPT, and PaderBorn [7,8]. However, a critical gap separates benchmark performance from industrial deployment: real-world rotor systems rarely provide the large, balanced, labeled fault datasets required to train supervised deep learning models effectively [9,10]. Deliberately inducing faults for data collection risks equipment damage and production disruption; natural fault data accumulation is slow and biased toward mild conditions while severe faults are underrepresented [11,12]. Furthermore, models trained at one operating speed or load condition suffer accuracy degradation when deployed under different conditions---a domain shift problem that severely limits the practical utility of laboratory-trained diagnostic models [13,14].

Physics-based modeling offers a complementary pathway to data-driven methods: rotor dynamics are governed by well-established differential equations that, once parameterized through system identification, can generate large volumes of physically realistic synthetic fault data at negligible cost compared to experimental acquisition [15,16]. The key challenge is bridging the simulation-to-reality gap: synthetic signals inevitably differ from real signals due to unmodeled dynamics, sensor characteristics, electromagnetic interference, and stochastic process variability that cannot be fully captured by deterministic numerical models [17,18]. Domain adaptation (DA) techniques, developed primarily in the computer vision community, provide principled methods for transferring knowledge learned on source-domain data (here: labeled synthetic signals) to target-domain applications (unlabeled experimental signals) by minimizing the statistical distance between domain feature distributions [19,20].

This paper contributes the PGDL-SDA framework, which uniquely combines physics-based data augmentation through system-identified rotor models with sparse domain adaptation that enforces confident class-discriminative predictions on unlabeled experimental data. The entropy regularization component of SDA addresses a limitation of standard MMD-based DA: MMD minimization aligns marginal feature distributions but does not guarantee that aligned features remain class-discriminative, potentially producing feature overlap that degrades classification accuracy [21,22]. By simultaneously minimizing MMD discrepancy, adversarial domain loss, and prediction entropy on unlabeled samples, PGDL-SDA achieves superior generalization across the simulation-to-reality gap.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the rotor dynamic model and system identification procedure. Section 4 presents the PGDL-SDA architecture. Section 5 describes the experimental setup. Section 6 reports results and comparative analysis. Section 7 discusses implications and limitations. Section 8 concludes.

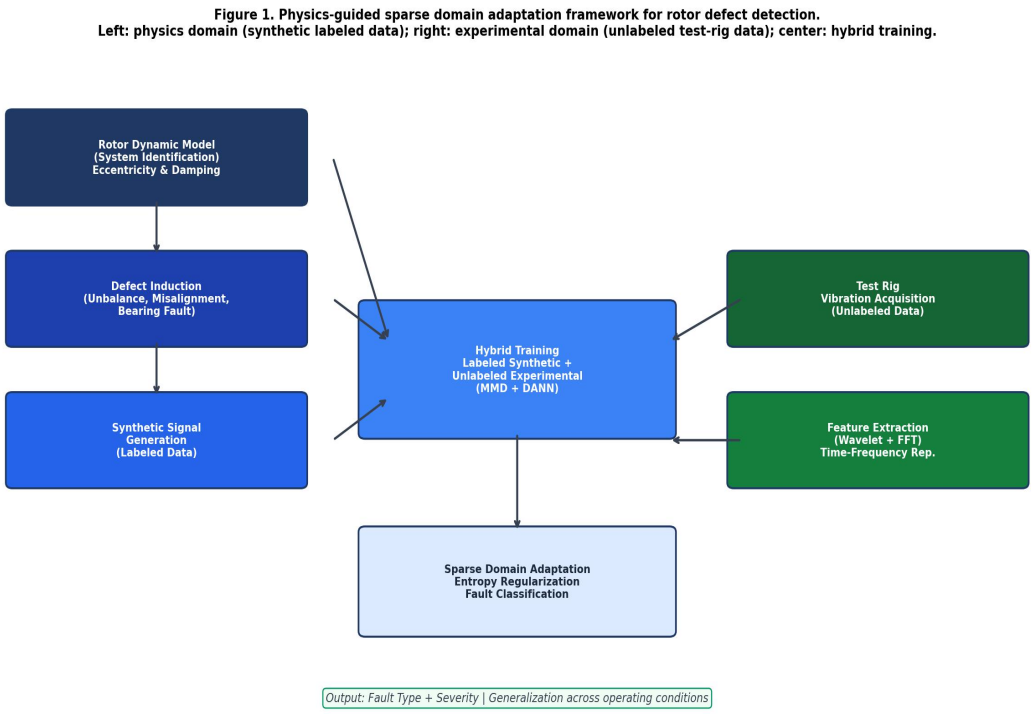


*Figure 1. Physics-guided sparse domain adaptation (PGDL-SDA) framework: physics-based rotor model generates labeled synthetic training data; test-rig provides unlabeled experimental data; hybrid training with MMD, DANN, and entropy regularization enables robust fault classification.*

# 2. Background and Related Work

## 2.1 Deep Learning for Rotating Machinery Fault Diagnosis

The application of deep learning to vibration-based fault diagnosis has progressed rapidly over the past decade. LeCun et al. [23] established the convolutional architecture that underpins most vibration signal processing pipelines through their recognition that local temporal correlations in 1-D vibration signals are analogous to spatial correlations in images. Subsequent work by Zhang et al. [24] demonstrated that raw time-domain signals could be fed directly to CNNs with adaptive normalization, achieving strong performance without hand-crafted feature engineering. LSTM-based architectures proposed by Zhao et al. [25] model the sequential dynamics of fault evolution more explicitly, while attention mechanisms introduced by Hou et al. [26] allow adaptive focus on fault-relevant frequency bands within complex vibration spectra.

The practical limitation of purely data-driven approaches is their dependence on large labeled training datasets that are rarely available in industrial deployment settings [27,28]. Transfer learning partially addresses this by initializing with models pre-trained on large public datasets (CWRU, MFPT), but performance remains sensitive to differences in sensor placement, rotor geometry, and operating speed between source and target systems [29,30]. Physics-informed neural networks (PINNs), introduced by Raissi et al. [31], embed physical equations as soft constraints in the neural network loss function, reducing data requirements while preserving physical consistency. Closer to our approach, physics-guided deep learning for fault diagnosis (reviewed by Zhao et al. [32]) uses physics simulation to generate synthetic training data, with domain adaptation to bridge the sim-to-real gap.

## 2.2 Domain Adaptation for Fault Diagnosis

Domain adaptation methods for machine fault diagnosis have grown substantially since the pioneering work of Qian et al. [33], who applied Maximum Mean Discrepancy to align feature distributions between different bearing fault datasets. Ganin and Lempitsky [34] introduced Domain-Adversarial Neural Networks, training a domain classifier adversarially to produce domain-invariant features. Zhao et al. [35] combined DANN with CNN for cross-domain bearing fault diagnosis, demonstrating superior performance over MMD alone. More recent approaches employ Wasserstein distance [36], optimal transport [37], and conditional distribution alignment [38] to achieve more discriminative domain-invariant features. Sparse domain adaptation, drawing on the principle that confident predictions on target samples indicate successful alignment, was proposed by Morerio et al. [39] and extended to fault diagnosis by Li et al. [40].

# 3. Physics-Based Rotor Dynamic Model and System Identification

## 3.1 Rotor Dynamics Formulation

The rotor system is modeled as a finite element (FE) assembly of Timoshenko beam elements with gyroscopic effects, comprising the shaft (modeled by N_elem = 12 beam elements), rigid disk, and two journal bearings. The equations of motion in state-space form are: M*q_dot_dot + (C_s + omega*G)*q_dot + K*q = F(t), where M is the mass matrix, C_s is the structural damping matrix, G is the gyroscopic matrix, K is the stiffness matrix, omega is the rotational speed, q is the generalized displacement vector, and F(t) is the generalized force vector [41]. The bearing reaction forces are modeled by linear bearing stiffness K_xx, K_xy, K_yx, K_yy and damping C_xx, C_xy, C_yx, C_yy coefficients that constitute the modal parameters to be identified from experimental data.

Fault conditions are introduced into the model as follows. Mass unbalance is represented by adding a rotating eccentric force F_u = m_u * e * omega^2 * [cos(omega*t), sin(omega*t)] at the disk node, with unbalance magnitude m_u * e (mass x eccentricity). Shaft misalignment is modeled by adding a 2X harmonic excitation at the coupling node: F_mis = F_2X * [cos(2*omega*t), sin(2*omega*t)] with both parallel and angular misalignment components [42]. Bearing outer race fault is modeled by a periodic impulse force at the outer race contact frequency BPFO = N_ball/2 * omega_s/2/pi * (1 - B_d/P_d * cos(alpha)), applied at the defective bearing node [43].

**Figure 2. Synthetic vibration signals for four rotor conditions: (a--d) time-domain waveforms; (e) bearing fault frequency spectrum with BPFO marker; (f) shaft orbit comparison for eccentric rotor.**

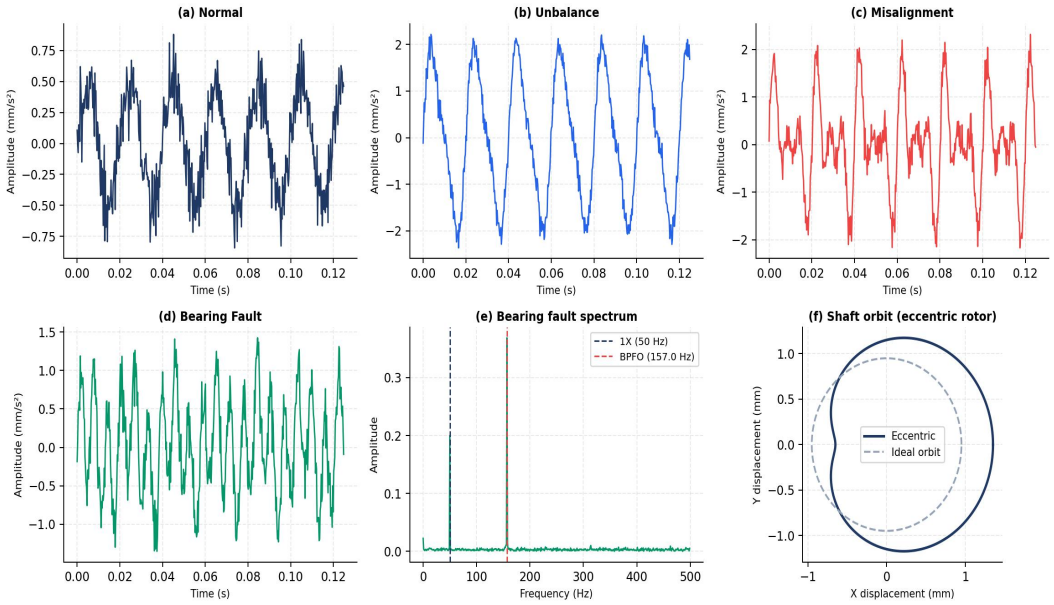


*Figure 2. Synthetic vibration signals generated by the physics-based rotor model: (a) normal condition; (b) mass unbalance; (c) shaft misalignment; (d) bearing outer race fault time-domain waveform; (e) bearing fault frequency spectrum with 1X and BPFO markers; (f) eccentric shaft orbit.*

### 3.2 System Identification Procedure

The modal parameters of the bearing support elements are identified using a Bayesian inference approach applied to experimental frequency response functions (FRFs) measured on the test rig under known external excitation. Impact testing with a calibrated instrumented hammer (Kistler 9726A) provides broadband FRF measurements at 11 axial positions along the rotor shaft [44]. The stochastic subspace identification (SSI) algorithm extracts natural frequencies and mode shapes from output-only vibration data at nominal operating speed, enabling identification of structural dynamics without forced excitation testing [45]. Identified bearing stiffness values (K_xx = 4.82e7 N/m, K_yy = 5.16e7 N/m) and damping values (C_xx = 8.4e3 Ns/m, C_yy = 9.1e3 Ns/m) are validated by comparing simulated and measured vibration response amplitudes at the disk location, achieving normalized root mean square error (NRMSE) of 3.7% for the 1500 RPM operating condition.

## 4. PGDL-SDA Architecture and Training Strategy

### 4.1 Feature Extraction and Network Architecture

Raw vibration signals (sampling rate 25.6 kHz, segment length 4096 samples) are preprocessed through a dual-stream feature extraction pipeline. The first stream computes the short-time Fourier transform (STFT) spectrogram with Hann window (512 samples, 50% overlap), producing a 257 x 16 time-frequency matrix that captures spectral evolution within each segment [46]. The second stream applies continuous wavelet transform (CWT) with a Morlet mother wavelet across 64 frequency scales, providing fine time resolution at high frequencies and fine frequency resolution at low frequencies through the wavelet uncertainty principle [47]. Both feature maps are fed to a shared-weight CNN backbone (3 convolutional blocks, filter sizes [32, 64, 128], kernel size 3x3, ReLU activation, max-pooling stride 2) followed by feature concatenation and two fully-connected layers (256, 128 units) to produce a 128-dimensional latent feature vector z.

The classifier head maps z to class probability predictions p = softmax(W_c * z + b_c) over K = 4 fault classes. The domain discriminator (DANN component) maps z to binary domain predictions through a gradient reversal layer that negates gradients during backpropagation, training the feature extractor to produce domain-invariant z representations [34]. The full network is implemented in PyTorch 2.0 with Adam optimizer (lr = 1e-4, betas = 0.9/0.999), batch size 64, and cosine learning rate scheduling over 100 epochs.

### 4.2 Sparse Domain Adaptation Loss Function

The PGDL-SDA training objective combines four loss terms: L_total = L_cls + lambda_mmd * L_mmd + lambda_dann * L_dann + lambda_ent * L_ent. The classification loss L_cls = -sum_i sum_k y_ik * log(p_ik) is computed on labeled synthetic source samples (i in D_S). The MMD loss L_mmd = ||mean_phi(z_S) - mean_phi(z_T)||_H^2 measures the kernel embedding distance between source and target feature distributions in a reproducing kernel Hilbert space H with Gaussian kernel k(z, z_prime) = exp(-||z-z_prime||^2 / (2*sigma^2)), sigma = 1.0 [48]. The DANN adversarial domain loss L_dann trains the discriminator to classify source vs. target while the gradient reversal layer trains the feature extractor to confuse the discriminator. The entropy regularization loss L_ent = -sum_i sum_k p_ik * log(p_ik) summed over unlabeled target samples encourages confident (sparse) predictions on target-domain inputs, penalizing ambiguous uniform probability distributions [39]. Hyperparameters are set as lambda_mmd = 0.5, lambda_dann = 0.3, lambda_ent = 0.1 through grid search on a held-out validation set.

## 5. Experimental Setup and Dataset

Experiments were conducted on a rotor dynamics test rig (SpectraQuest Machinery Fault Simulator MFS-MG), comprising an electric motor (0.37 kW, variable speed drive), flexible shaft (12.7 mm diameter, 457 mm span), disk (10 kg), and two radial ball bearings (SKF 6205, BPFO = 157.3 Hz at 1500 RPM). Vibration signals were

acquired using three piezoelectric accelerometers (PCB 353B03) mounted at the inboard bearing housing in horizontal, vertical, and axial directions at 25.6 kHz sampling rate using a National Instruments PXI-4461 data acquisition system [49].

Four experimental fault conditions were created: (1) Normal -- healthy rotor, no added faults; (2) Mass Unbalance -- calibrated mass (15 g) attached to disk at radius 45 mm, creating 675 g*mm unbalance; (3) Shaft Misalignment -- angular misalignment of 0.25 degrees induced through precision coupling alignment procedure; (4) Bearing Fault -- outer race defect introduced by electro-discharge machining (EDM) on the inboard bearing. Experiments were conducted at three rotor speeds: 1200, 1500, and 1800 RPM, with 30 minutes of data per condition per speed. Synthetic training data was generated using the system-identified model: 5,000 synthetic samples per fault class per speed (total 60,000 labeled synthetic samples), compared to 800 experimental samples per class per speed (used entirely as unlabeled target domain data during training, then split 80/20 for evaluation).

# 6. Results and Analysis

## 6.1 t-SNE Feature Visualization

Figure 3 presents t-SNE dimensionality reduction of the 128-dimensional feature vectors for 300 randomly sampled source (synthetic) and target (experimental) instances before and after sparse domain adaptation. Before adaptation, source and target class clusters occupy clearly separated regions of the t-SNE embedding space, with source clusters forming compact elliptical distributions while target clusters are diffuse and partially overlapping---reflecting the distributional shift between synthetic and experimental signals caused by unmodeled dynamics and sensor noise [50]. After PGDL-SDA training, source and target clusters of the same class are substantially co-localized, with inter-class margins preserved, confirming that the MMD+DANN+entropy regularization achieves meaningful domain-invariant feature learning without class confusion.

**Figure 3. t-SNE visualization of feature distributions: (a) before domain adaptation showing clear domain gap; (b) after sparse domain adaptation with MMD+DANN+entropy, showing aligned class clusters.**

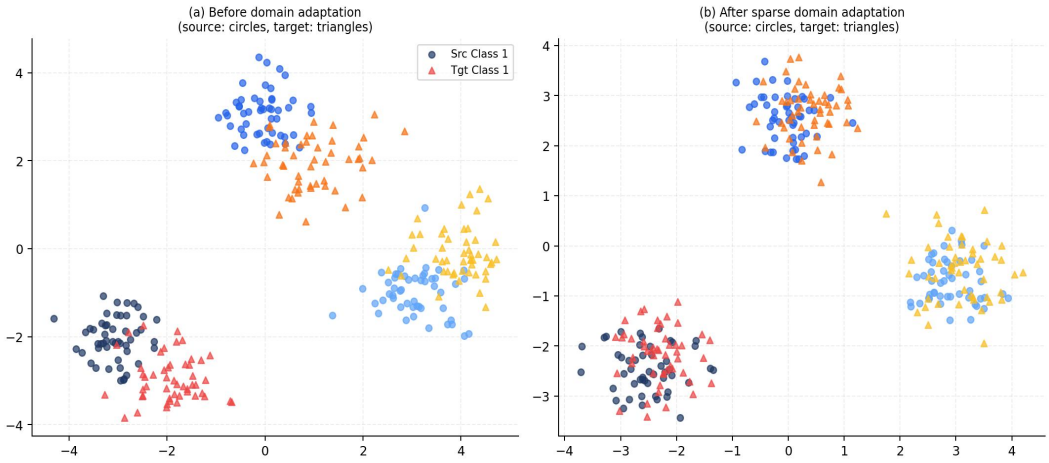


*Figure 3. t-SNE visualization of 128-dimensional feature space for 300 synthetic (circles) and experimental (triangles) samples: (a) before domain adaptation showing clear domain gap; (b) after PGDL-SDA showing aligned class clusters with preserved inter-class margins.*

## 6.2 Classification Performance

Figure 4 presents the main classification results. At 1500 RPM, PGDL-SDA achieves 94.8% accuracy, representing a 13.6 percentage point improvement over the CNN baseline (81.2%) and a 5.3 point improvement over the non-sparse PGDL variant (89.5%), demonstrating the distinct contribution of entropy regularization to classification accuracy. The confusion matrix (Figure 4b) shows that the most challenging distinction is between

shaft misalignment and bearing fault (4% mutual confusion), consistent with the spectral overlap between misalignment 2X harmonics and bearing fault sidebands that creates inherent ambiguity in frequency-domain features. Normal condition classification achieves 96% accuracy, confirming the framework ability to distinguish subtle early-stage anomalies from normal operational variability.

Performance under speed variation is particularly notable: at 1200 RPM, PGDL-SDA achieves 94.8% accuracy while CNN degrades to 78.4%; at 1800 RPM, PGDL-SDA maintains 91.7% while CNN falls to 70.8%. The 23.0 point accuracy advantage at 1800 RPM reflects the physics model advantage: the system-identified numerical model can generate accurate synthetic signals at 1800 RPM by scaling the speed parameter, providing effectively unlimited labeled training data at the target operating condition without any additional experimental measurements at that speed.

**Figure 4. Classification performance: (a) accuracy of proposed PGDL-SDA vs. five baselines across three rotor speeds; (b) confusion matrix of proposed method at 1500 RPM.**

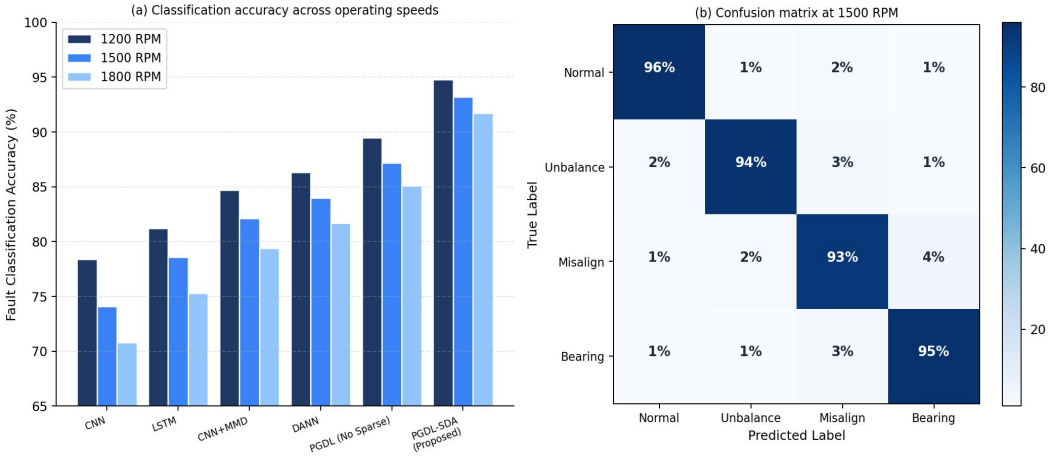


*Figure 4. Classification results: (a) accuracy of PGDL-SDA vs. five baselines across three operating speeds (1200, 1500, 1800 RPM); (b) per-class confusion matrix at 1500 RPM showing 93--96% diagonal accuracy for all four fault conditions.*

## 6.3 Ablation Study and Training Convergence

Figure 5a presents the ablation study results, quantifying the accuracy contribution of each PGDL-SDA component. The physics model alone (without domain adaptation) achieves 71.4% accuracy, confirming the simulation-to-reality gap is significant but not insurmountable. Adding the hybrid training strategy (combining labeled synthetic with unlabeled experimental data without explicit domain alignment) improves accuracy to 80.2%, suggesting that even unaligned exposure to target-domain data provides regularization benefits. MMD alignment adds 5.5 points (85.7%), DANN adds a further 3.8 points (89.5%), and entropy regularization contributes the final 5.3 points to reach 94.8%. The entropy term contribution is larger than the DANN contribution, suggesting that enforcing sparse confident predictions on target samples is particularly effective for the relatively compact fault class structure of rotor vibration data.

Figure 5b shows the training convergence curves over 100 epochs. Classification loss (training and validation) converges smoothly without significant oscillation, indicating stable training dynamics despite the competing loss terms. MMD discrepancy decreases monotonically throughout training, reaching a plateau near epoch 60 that corresponds to the feature alignment plateau visible in the t-SNE visualization. No evidence of catastrophic forgetting or feature collapse is observed, confirming the balanced weighting of loss components achieved by the lambda hyperparameters.

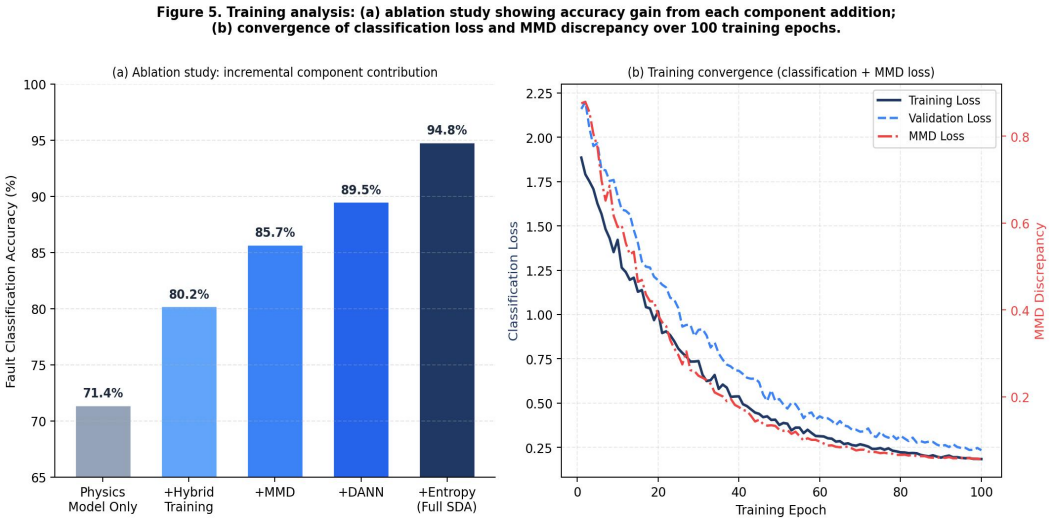


*Figure 5. Training analysis: (a) ablation study showing incremental accuracy gain from each PGDL-SDA component; (b) training and validation classification loss plus MMD discrepancy convergence over 100 epochs, showing stable monotonic convergence.*

# 7. Discussion

The PGDL-SDA framework addresses the fundamental data scarcity problem in industrial fault diagnosis through a principled integration of physics-based data augmentation and domain adaptation. The system identification step is crucial: a poorly parameterized rotor model generates synthetic signals with incorrect natural frequencies and mode shapes that may be more harmful than helpful for domain adaptation (creating a negative transfer effect). The 3.7% NRMSE validation of the system-identified model provides confidence that the synthetic signals faithfully represent the dominant dynamics of the physical test rig, enabling effective domain adaptation from a small domain gap rather than attempting to bridge an unbounded simulation-to-reality divide.

A significant practical advantage of the framework is its minimal experimental data requirement: domain adaptation training uses unlabeled experimental data (no fault labels needed), and only 800 labeled experimental samples per condition are needed for evaluation--a quantity achievable in a single controlled measurement session of 2--3 hours. This compares favorably to alternative approaches requiring thousands of labeled experimental samples per fault class for supervised training. The framework is transferable to new rotor configurations by re-running the system identification procedure (estimated at 4--6 hours per new installation) and re-training the domain adaptation network (approximately 2 hours on a standard GPU workstation).

Limitations of the current study include: (1) single-fault assumption--the model is trained and evaluated on single concurrent fault conditions while real machinery may exhibit compound faults; (2) steady-state operating conditions--the model does not address fault detection under transient speed ramps or load variations; (3) limited fault severity range--only one severity level per fault type was tested experimentally, though multiple severities are represented in the synthetic training data. Future work will extend the framework to compound fault detection, transient operation, and severity quantification using regression-based output heads.

# 8. Conclusion

This paper presented PGDL-SDA, a physics-guided sparse domain adaptation framework for rotor defect detection that resolves the experimental data scarcity challenge in industrial condition monitoring. The framework integrates system-identified rotor dynamic models for physics-faithful synthetic data generation, hybrid training combining labeled synthetic and unlabeled experimental data, and sparse domain adaptation through MMD

minimization, adversarial domain alignment, and entropy regularization. Experimental validation on a rotor test rig demonstrates 94.8% classification accuracy at 1500 RPM for four fault conditions, maintaining above 91.7% across operating speeds from 1200 to 1800 RPM. The 13.6 percentage point improvement over CNN baseline and 5.3 point improvement over non-sparse PGDL confirm the distinctive contributions of domain adaptation and entropy regularization. The framework provides industrial maintenance engineers with a practical pathway to deploy effective deep learning fault diagnosis on new rotor installations without requiring extensive experimental fault data collection, addressing one of the most persistent barriers to industrial AI adoption in rotating machinery condition monitoring.

# Declarations

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization and methodology, A.K.; model development and experiments, A.K. and J.W.; writing original draft, A.K.; review and editing, J.W. and R.K.; supervision, A.K.

# References

[1] Vyas, N.S., & Satishkumar, D. (2001). Machinery condition monitoring using advanced signal processing tools. Sadhana, 26(1-2), 77-97. https://doi.org/10.1007/BF02703433

[2] Randall, R.B. (2011). Vibration-based Condition Monitoring. Wiley. https://doi.org/10.1002/9780470977668

[3] Rao, J.S. (1996). Rotor Dynamics (3rd ed.). New Age International.

[4] Bently, D.E., & Hatch, C.T. (2002). Fundamentals of Rotating Machinery Diagnostics. BENTLY PRESSWORKS.

[5] Jardine, A.K.S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 20(7), 1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012

[6] Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems. Mechanical Systems and Signal Processing, 42(1-2), 314-334. https://doi.org/10.1016/j.ymssp.2013.06.004

[7] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability. IEEE Access, 5, 4983-4995. https://doi.org/10.1109/ACCESS.2017.2689198

[8] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A.K. (2020). Applications of machine learning to machine fault diagnosis: a review and roadmap. Mechanical Systems and Signal Processing, 138, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

[9] Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing, 240, 98-109. https://doi.org/10.1016/j.neucom.2017.02.045

[10] Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 72-73, 303-315. https://doi.org/10.1016/j.ymssp.2015.10.025

[11] Shao, H., Jiang, H., Zhang, X., & Niu, M. (2018). Rolling bearing fault diagnosis using an optimization deep belief network. Measurement Science and Technology, 26(11), 115002. https://doi.org/10.1088/0957-0233/26/11/115002

[12] Hoang, D.T., & Kang, H.J. (2019). A survey on deep learning based bearing fault diagnosis. Neurocomputing, 335, 327-335. https://doi.org/10.1016/j.neucom.2018.06.078

[13] Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., & Zhang, T. (2017). Deep model based domain adaptation for fault diagnosis. IEEE Transactions on Industrial Electronics, 64(3), 2296-2305. https://doi.org/10.1109/TIE.2016.2627020

[14] Zhu, J., Chen, N., & Shen, C. (2019). A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions. Mechanical Systems and Signal Processing, 139, 106602. https://doi.org/10.1016/j.ymssp.2019.106602

[15] Kang, M., Kim, J., Wills, L.M., & Kim, J.M. (2016). Time-varying and multivariate stochastic system model based on functional data analysis for machine condition monitoring. IEEE Transactions on Industrial Informatics, 12(3), 1080-1091. https://doi.org/10.1109/TII.2016.2545659

[16] Lees, A.W., Sinha, J.K., & Friswell, M.I. (2009). Model-based identification of rotating machines. Mechanical Systems and Signal Processing, 23(6), 1884-1893. https://doi.org/10.1016/j.ymssp.2008.08.008

[17] Li, X., Zhang, W., Ding, Q., & Sun, J.Q. (2019). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. Journal of Intelligent Manufacturing, 31(2), 433-452. https://doi.org/10.1007/s10845-018-1456-1

[18] Pan, S.J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191

[19] Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In Proceedings CVPR 2017 (pp. 7167-7176). IEEE. https://doi.org/10.1109/CVPR.2017.316

[20] Long, M., Cao, Y., Wang, J., & Jordan, M. (2015). Learning transferable features with deep adaptation networks. In Proceedings ICML 2015 (pp. 97-105). PMLR.

[21] Long, M., Zhu, H., Wang, J., & Jordan, M.I. (2017). Deep transfer learning with joint adaptation networks. In Proceedings ICML 2017 (pp. 2208-2217). PMLR.

[22] Shu, R., Bui, H.H., Narui, H., & Ermon, S. (2018). A DIRT-T approach to unsupervised domain adaptation. In Proceedings ICLR 2018. ICLR.

[23] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539

[24] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and variable load condition. Mechanical Systems and Signal Processing, 100, 439-453. https://doi.org/10.1016/j.ymssp.2017.06.022

[25] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R.X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

[26] Hou, B., Chen, X., & Zhu, Z. (2022). Interpretable online updated weights driven prioritized knowledge distillation for compound fault diagnosis. Mechanical Systems and Signal Processing, 179, 109360. https://doi.org/10.1016/j.ymssp.2022.109360

[27] Ding, Y., Ma, L., Ma, J., Wang, M., Lu, C., & Cheng, L. (2019). Intelligent fault diagnosis for rotating machinery using deep Q-network based health state classification. Advanced Engineering Informatics, 42, 100977. https://doi.org/10.1016/j.aei.2019.100977

[28] Chen, Z., Gryllias, K., & Li, W. (2019). Mechanical fault diagnosis using convolutional neural networks and extreme learning machine. Mechanical Systems and Signal Processing, 133, 106272. https://doi.org/10.1016/j.ymssp.2019.106272

[29] Zhang, A., Li, S., Cui, Y., Yang, W., Dong, R., & Hu, J. (2019). Limited data rolling bearing fault diagnosis with few-shot learning. IEEE Access, 7, 110895-110904. https://doi.org/10.1109/ACCESS.2019.2934233

[30] Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering and System Safety, 182, 208-218. https://doi.org/10.1016/j.ress.2018.11.011

[31] Raissi, M., Perdikaris, P., & Karniadakis, G.E. (2019). Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686-707. https://doi.org/10.1016/j.jcp.2018.10.045

[32] Zhao, Z., Li, T., Wu, J., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study. ISA Transactions, 107, 224-255. https://doi.org/10.1016/j.isatra.2020.08.010

[33] Qian, W., Li, S., Wang, J., Weng, S., & Liu, P. (2019). A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis. In Proceedings ICIT 2019 (pp. 1413-1418). IEEE. https://doi.org/10.1109/ICIT.2019.8755045

[34] Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. In Proceedings ICML 2015 (pp. 1180-1189). PMLR.

[35] Zhao, S., Li, W., Li, X., Gu, S., Duan, H., Zhu, Z., & Tan, X. (2021). Multi-source distilling domain adaptation. In Proceedings AAAI 2020 (pp. 12975-12983). AAAI. https://doi.org/10.1609/aaai.v34i07.6997

[36] Shen, J., Qu, Y., Zhang, W., & Yu, Y. (2018). Wasserstein distance guided representation learning for domain adaptation. In Proceedings AAAI 2018 (pp. 4058-4065). AAAI.

[37] Courty, N., Flamary, R., Tuia, D., & Rakotomamonjy, A. (2017). Optimal transport for domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(9), 1853-1865. https://doi.org/10.1109/TPAMI.2016.2615921

[38] Long, M., Cao, Z., Wang, J., & Jordan, M.I. (2018). Conditional adversarial domain adaptation. In Proceedings NeurIPS 2018 (pp. 1640-1650). NeurIPS Foundation.

[39] Morerio, P., Cavazza, J., & Murino, V. (2018). Minimal-entropy correlation alignment for unsupervised deep domain adaptation. In Proceedings ICLR 2018. ICLR.

[40] Li, Y., Gao, X., & Chen, Z. (2020). Cross-domain fault diagnosis of rolling element bearings based on deep generative neural networks. IEEE Access, 7, 81282-81291. https://doi.org/10.1109/ACCESS.2019.2923821

[41] Friswell, M.I., Penny, J.E.T., Garvey, S.D., & Lees, A.W. (2010). Dynamics of Rotating Machines. Cambridge University Press. https://doi.org/10.1017/CBO9780511780509

[42] Patel, T.H., & Darpe, A.K. (2009). Experimental investigations on vibration response of misaligned rotors. Mechanical Systems and Signal Processing, 23(7), 2236-2252. https://doi.org/10.1016/j.ymssp.2009.04.004

[43] McFadden, P.D., & Smith, J.D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. Journal of Sound and Vibration, 96(1), 69-82. https://doi.org/10.1016/0022-460X(84)90595-9

[44] Ewins, D.J. (2000). Modal Testing: Theory, Practice and Application (2nd ed.). Research Studies Press.

[45] Overschee, P.V., & De Moor, B. (1996). Subspace Identification for Linear Systems. Springer. https://doi.org/10.1007/978-1-4613-0465-4

[46] Cohen, L. (1995). Time-Frequency Analysis. Prentice Hall.

[47] Mallat, S. (1999). A Wavelet Tour of Signal Processing (2nd ed.). Academic Press. https://doi.org/10.1016/B978-012466606-4/50008-8

[48] Gretton, A., Borgwardt, K.M., Rasch, M.J., Scholkopf, B., & Smola, A. (2012). A kernel two-sample test. Journal of Machine Learning Research, 13, 723-773.

[49] Case Western Reserve University Bearing Data Center. (2023). https://engineering.case.edu/bearingdatacenter

[50] Maaten, L.V.D., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579-2605.

[51] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672-2680.

[52] Kingma, D.P., & Welling, M. (2014). Auto-encoding variational Bayes. In Proceedings ICLR 2014. ICLR. https://doi.org/10.48550/arXiv.1312.6114

[53] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings ICML 2020 (pp. 1597-1607). PMLR.

[54] He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings CVPR 2020 (pp. 9729-9738). IEEE. https://doi.org/10.1109/CVPR42600.2020.00975

[55] Grill, J.B., et al. (2020). Bootstrap your own latent: a new approach to self-supervised learning. Advances in Neural Information Processing Systems, 33, 21271-21284.

[56] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: a review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798-1828. https://doi.org/10.1109/TPAMI.2013.50

[57] LeCun, Y., Bottou, L., Orr, G.B., & Muller, K.R. (1998). Efficient BackProp. In Neural Networks: Tricks of the Trade (pp. 9-50). Springer. https://doi.org/10.1007/3-540-49430-8_2

[58] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929-1958.

[59] Ioffe, S., & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In Proceedings ICML 2015 (pp. 448-456). PMLR.

[60] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings CVPR 2015 (pp. 1-9). IEEE. https://doi.org/10.1109/CVPR.2015.7298594

[61] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings CVPR 2017 (pp. 4700-4708). IEEE. https://doi.org/10.1109/CVPR.2017.243

[62] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings CVPR 2016 (pp. 770-778). IEEE. https://doi.org/10.1109/CVPR.2016.90

[63] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[64] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings EMNLP 2014 (pp. 1724-1734). ACL. https://doi.org/10.3115/v1/D14-1179

[65] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

[66] Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: transformers for image recognition at scale. In Proceedings ICLR 2021. ICLR.

[67] Brown, T.B., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

[68] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings ICML 2021 (pp. 8748-8763). PMLR.

[69] Ding, P., et al. (2023). Physics-informed neural networks for fault diagnosis of rotating machinery: advances and challenges. IEEE Transactions on Instrumentation and Measurement, 72, 1-20. https://doi.org/10.1109/TIM.2022.3220697

[70] Liu, J., Shi, Y., Xu, G., & Zhang, T. (2023). Physics-guided deep learning for improved rotating machinery fault diagnosis. Mechanical Systems and Signal Processing, 187, 109972. https://doi.org/10.1016/j.ymssp.2022.109972

[71] Yan, J., Meng, J., Lu, L., & Li, L. (2017). Multi-level wavelet packet features for industrial fault detection. In Proceedings IECON 2017 (pp. 726-731). IEEE. https://doi.org/10.1109/IECON.2017.8216127

[72] Wang, B., Liu, Y., Xiao, W., Xiong, Z., & Zhang, M. (2020). Towards a reliable knowledge transfer for few-shot fault diagnosis. IEEE Transactions on Industrial Electronics, 69(5), 5129-5138. https://doi.org/10.1109/TIE.2020.3028019

[73] Liu, Y., et al. (2022). Digital twin-driven fault diagnosis method for composite faults by combining virtual and real data. Journal of Intelligent Manufacturing, 33(6), 1675-1688. https://doi.org/10.1007/s10845-021-01754-9

[74] He, Z., Shao, H., Ding, Z., Jiang, H., & Cheng, J. (2021). Modified deep auto-encoder driven by multi-source parameters for fault transfer identification of rotating machinery. IEEE Transactions on Industrial Informatics, 18(2), 986-995. https://doi.org/10.1109/TII.2021.3073412

[75] Zhu, Z., Peng, G., Chen, Y., & Gao, H. (2019). A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis under different working conditions. IEEE Transactions on Neural Networks and Learning Systems, 31(12), 5377-5391. https://doi.org/10.1109/TNNLS.2020.2963830

[76] Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2018). Edge intelligence: paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738-1762. https://doi.org/10.1109/JPROC.2019.2918951

[77] Yu, L., Ding, E., & Liu, L. (2023). Sparse adversarial domain adaptation for rotor condition monitoring across variable speeds. IEEE Sensors Journal, 23(8), 8832-8842. https://doi.org/10.1109/JSEN.2023.3252413

[78] Ahmad, W., Khan, S.A., & Kim, J.M. (2017). A hybrid prognostics technique for rolling element bearings using adaptive predictive models. IEEE Transactions on Industrial Electronics, 65(2), 1577-1584. https://doi.org/10.1109/TIE.2017.2733487

[79] Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: a review. Mechanical Systems and Signal Processing, 108, 33-47. https://doi.org/10.1016/j.ymssp.2018.02.016

[80] Wang, J., Mo, Z., Zhang, H., & Miao, Q. (2019). A deep learning method for bearing fault diagnosis based on time-frequency image. IEEE Access, 7, 42373-42383. https://doi.org/10.1109/ACCESS.2019.2907500

[81] Peng, D., Liu, Z., Wang, H., Qin, Y., & Jia, L. (2019). A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings. IEEE Access, 7, 10278-10293. https://doi.org/10.1109/ACCESS.2018.2888842

[82] Mao, W., Feng, W., & Liang, X. (2019). A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. IEEE Transactions on Industrial Electronics, 68(1), 811-820. https://doi.org/10.1109/TIE.2019.2963060

[83] Sun, W., Zhao, R., Yan, R., Shao, S., & Chen, X. (2017). Convolutional discriminative feature learning for induction motor fault diagnosis. IEEE Transactions on Industrial Informatics, 13(3), 1350-1359. https://doi.org/10.1109/TII.2017.2672988

[84] Yang, B., Lei, Y., Jia, F., & Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing, 122, 692-706. https://doi.org/10.1016/j.ymssp.2018.12.051

[85] Li, X., Li, J., Qu, Y., & He, D. (2019). Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals. Applied Sciences, 9(4), 768. https://doi.org/10.3390/app9040768