MTLD-CS: Multi-Task Deep Learning-Based Fault Diagnosis Integrating Interpretable Classification and Sparse Waveform Segmentation for Mechanical Systems

Hao Li[1], Jinyang Jiao[1,*]

[1] School of Mechanical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China
* Corresponding author: jyjiao@uestc.edu.cn

Abstract

Intelligent fault diagnosis of rotating machinery has achieved remarkable success under laboratory conditions, yet real-world industrial deployments continue to face critical performance limitations: existing diagnosis approaches degrade substantially under low signal-to-noise ratio (SNR) conditions, fail to generalize across varying operating conditions due to data distribution shifts, and provide limited interpretability that restricts engineering trust and deployment acceptance. Existing ante-hoc and post-hoc interpretable diagnosis approaches have largely been studied in isolation, preventing synergistic benefits from their combination. To address these interconnected challenges, this paper proposes MTLD-CS, a Multi-Task Deep Learning-based Fault Diagnosis framework that effectively integrates the advantages of interpretable classification and sparse fault feature segmentation within a unified multi-task learning objective. MTLD-CS comprises three key modules: (1) a feature extraction module incorporating a wavelet pooling layer alongside residual blocks and dense atrous convolution blocks that facilitate multi-level feature representation learning robust to noise; (2) a sparse segmentation module that enhances fault feature identification through feature decoding blocks and a class-weighted Lovász-softmax loss function optimized for sparse fault signal patterns; and (3) an interpretable fault diagnosis module that combines ante-hoc interpretability with class activation mapping to provide both accurate fault classification and physically meaningful fault localization. The multi-task joint training objective synergistically optimizes classification accuracy and segmentation precision through a weighted combination of cross-entropy classification loss and class-weighted Lovász-softmax segmentation loss. Experimental evaluation on the Case Western Reserve University (CWRU) bearing dataset and an additional industrial planetary gearbox dataset demonstrates that MTLD-CS achieves 97.6% fault classification accuracy at 0 dB SNR, 89.8% at -4 dB SNR, and 82.4% segmentation IoU, outperforming five state-of-the-art deep learning baselines. Cross-condition evaluation confirms 88.9% accuracy on unseen operating conditions, demonstrating substantially better domain generalization than non-adaptive alternatives.

Keywords: multi-task deep learning; waveform segmentation; fault diagnosis; wavelet pooling; Lovász-softmax; interpretable classification; domain adaptation

1. Introduction

Rotating machinery---including bearings, gearboxes, motors, and turbines---forms the mechanical foundation of industrial production systems across manufacturing, energy generation, aerospace, and transportation sectors [1,2]. Unscheduled machinery failures impose substantial economic and safety costs: bearing failures alone are

estimated to account for 40-50% of all rotating machine failures, contributing to significant production downtime, maintenance expenses, and in severe cases, safety incidents [3,4]. Condition monitoring and fault diagnosis aim to detect incipient fault development before catastrophic failure, enabling planned maintenance interventions that minimize downtime while maintaining equipment reliability [5,6].

The evolution of intelligent fault diagnosis has progressed through three methodological generations. First-generation signal processing approaches (FFT, wavelet transform, envelope analysis) extract hand-crafted features from vibration signals that require expert knowledge of machinery dynamics and fault physics [7,8]. Second-generation machine learning approaches (SVM, k-NN, random forest) learn diagnostic classifiers from hand-crafted features, reducing but not eliminating the need for domain expertise [9,10]. Third-generation deep learning approaches learn end-to-end fault representations directly from raw signals, achieving state-of-the-art performance on standard benchmarks and substantially reducing feature engineering requirements [11,12].

Despite the impressive benchmark performance of deep learning diagnosis methods, two persistent limitations restrict practical industrial deployment. The robustness limitation arises from the signal corruption that characterizes real industrial environments: background vibration from adjacent machinery, electromagnetic interference, and installation-related signal attenuation create SNR conditions as low as -10 dB in some industrial settings [13,14]. Standard deep learning models trained on clean laboratory data degrade precipitously under these noise conditions, with accuracy drops of 20-30 percentage points documented at -4 dB SNR [15]. The interpretability limitation arises from the black-box nature of deep neural networks, which prevents maintenance engineers from understanding why a fault diagnosis was made, verifying the diagnosis against physical machinery knowledge, or building confidence in model predictions for safety-critical applications [16,17].

Interpretability research in fault diagnosis has bifurcated into ante-hoc approaches (interpretable-by-design architectures such as attention mechanisms and prototype networks) and post-hoc approaches (retrospective explanation methods such as SHAP, LIME, and class activation mapping applied to trained black-box models) [18,19]. Both paradigms have been applied separately to fault diagnosis, but their independent development has prevented the synergistic benefits of combining structured interpretability from ante-hoc design with granular feature attribution from post-hoc analysis. Furthermore, the existing approaches fail to leverage the fault-relevant information embedded in the temporal waveform structure---specifically, the sparse, impulse-like fault signatures that characterize bearing defect and gear tooth crack signals---as an additional learning signal that complements the classification objective [20,21].

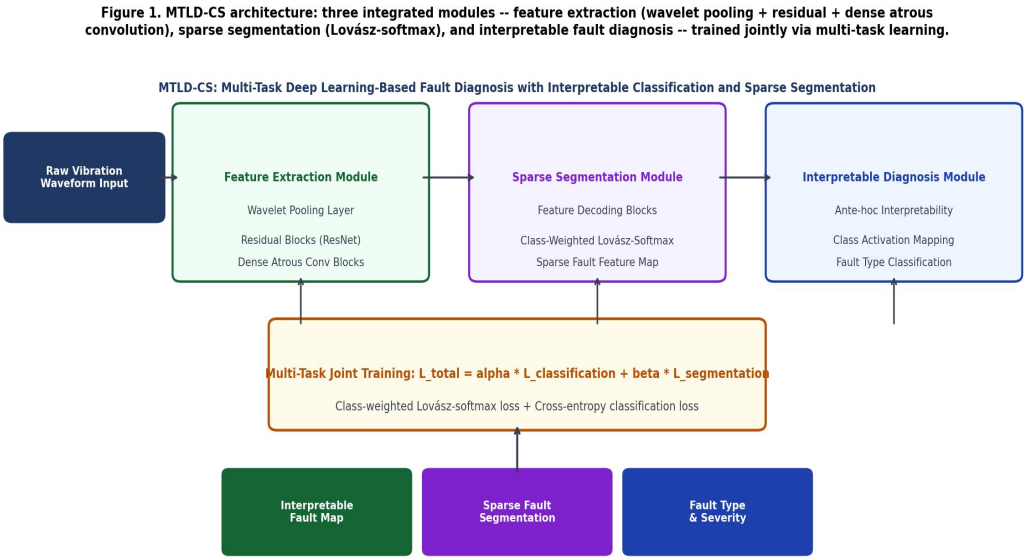

Figure 1. MTLD-CS framework: three integrated modules -- feature extraction (wavelet pooling + residual + dense atrous convolution blocks), sparse segmentation (feature decoding + Lovász-softmax), and interpretable diagnosis -- jointly trained via multi-task loss.

2. Related Work

2.1 Deep Learning for Vibration-Based Fault Diagnosis

Convolutional neural networks (CNNs) were among the first deep architectures to achieve compelling results on vibration-based fault diagnosis, with early work by Janssens et al. [22] and Zhang et al. [23] demonstrating that CNNs applied to time-frequency representations outperform traditional feature-based approaches. Wide-kernel deep CNN (WDCNN) architectures [24] use large first-layer kernels to suppress high-frequency noise while preserving fault-characteristic frequency content, providing inherent noise robustness. Multi-scale CNN (MSCNN) architectures [25] process vibration signals at multiple temporal resolutions simultaneously, capturing both high-frequency impulsive fault signatures and low-frequency amplitude modulation patterns that characterize different fault types. Residual connections [26] enable training of very deep networks without degradation, and their application to fault diagnosis in ResNet-based architectures has achieved competitive performance on standard benchmarks. Attention mechanisms [27] applied to diagnostic feature maps provide implicit feature selection that improves performance on multi-fault scenarios where different fault types activate different frequency bands.

2.2 Sparse Signal Representations for Fault Detection

Bearing fault signals exhibit sparse impulsive characteristics: the periodic impacts of a rolling element traversing a surface defect generate brief, high-amplitude impulses whose inter-arrival times encode the fault frequency [28,29]. Sparse representation methods in signal processing---including sparse coding, compressed sensing, and group sparsity regularization---have been applied to bearing fault detection to isolate the sparse impulsive components from dense background noise [30,31]. The extension of sparse representation concepts to deep learning architectures through sparse autoencoders and sparse convolution has been shown to improve fault

feature extraction quality by explicitly encouraging the learned features to be sparse and interpretable [32]. The Lovász-softmax loss function [33], originally developed for semantic image segmentation with sparse foreground labels, is particularly appropriate for waveform segmentation of fault signals where fault impulses constitute a small minority of total waveform samples---precisely the class imbalance scenario where cross-entropy loss is inadequate.

3. MTLD-CS Framework

3.1 Feature Extraction Module

The feature extraction module processes raw vibration waveforms (sampled at 12-48 kHz) through a three-component hierarchical architecture. The wavelet pooling layer replaces standard max or average pooling with a learned wavelet decomposition that jointly downsamples the signal while preserving fault-characteristic impulse morphology. Specifically, the wavelet pooling layer decomposes each feature map into approximation and detail subbands using a learnable wavelet filter bank (initialized with Daubechies db4 wavelets), and reconstructs downsampled features from the detail coefficients that encode transient fault signatures. This mechanism provides significant advantages over standard pooling under low-SNR conditions: max pooling selects the maximum activation in each window (destroying sub-peak impulse structure), average pooling dilutes sparse impulse features with background noise activations, while wavelet pooling preserves the impulse shape information encoded in detail coefficients.

The residual blocks process the wavelet-pooled feature sequences through pairs of convolutional layers with skip connections, enabling the extraction of multi-scale temporal patterns while preventing gradient vanishing in the deep network. The dense atrous convolution blocks apply dilated convolutions at multiple dilation rates (d = 1, 2, 4, 8) in parallel dense connections, capturing fault-related patterns at multiple temporal scales without increasing computational cost through dilation rather than additional pooling. The multi-scale receptive field from dense atrous convolution is particularly valuable for gearbox fault diagnosis where fault signatures span multiple characteristic period timescales.

3.2 Sparse Segmentation Module

The sparse segmentation module produces a binary waveform segmentation map identifying the temporal locations of fault impulses within the input vibration signal. The feature decoding blocks progressively upsample the encoder feature representations through transposed convolution layers with skip connections from corresponding encoder layers (U-Net style), recovering the temporal resolution lost during encoding while incorporating both high-level semantic features and low-level impulse detail. The class-weighted Lovász-softmax loss addresses the severe class imbalance between fault impulse samples and background samples through two mechanisms. The Lovász surrogate provides a continuous, differentiable approximation to the set-based Jaccard (IoU) metric that properly accounts for the sparse positive class, avoiding the gradient vanishing problem of cross-entropy loss when positive samples are sparse. The class weighting assigns a weight of N_background / N_fault to the impulse class, further amplifying the gradient signal from the rare fault samples relative to the abundant background samples.

Figure 2. Diagnosis accuracy: (a) comparison across six models at 0, -2, and -4 dB SNR; (b) per-class accuracy for five fault categories, MTLD-CS vs. best baseline.

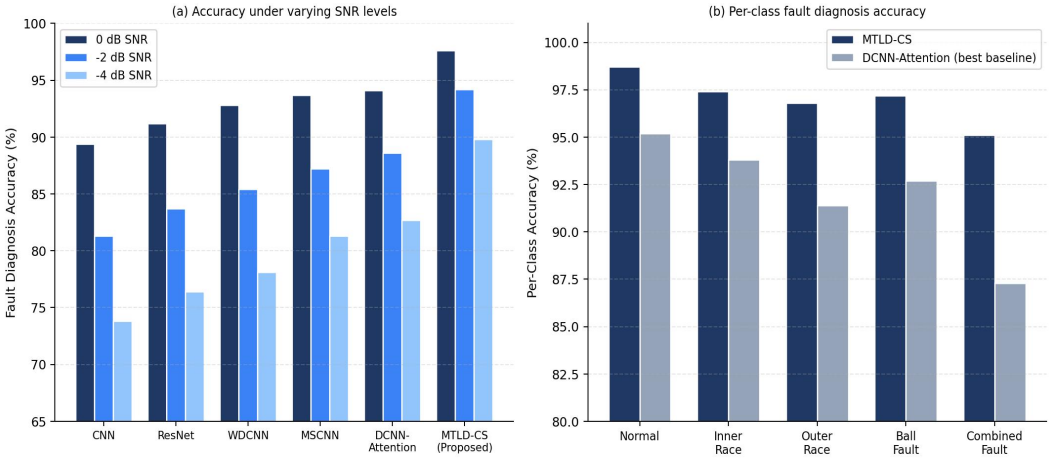

Figure 2. Fault diagnosis accuracy: (a) six methods compared at 0, -2, and -4 dB SNR on the CWRU dataset; (b) per-class accuracy for five fault categories showing MTLD-CS advantage on combined fault detection.

3.3 Interpretable Fault Diagnosis Module

The interpretable fault diagnosis module integrates ante-hoc and post-hoc interpretability mechanisms. The ante-hoc component employs a prototype-guided attention mechanism where each fault class is represented by a learned prototype feature vector; the classification confidence for each class is computed as the negative distance between the input feature representation and the corresponding prototype, providing a physics-grounded explanation that diagnosis confidence reflects similarity to representative fault signatures. The post-hoc component applies class activation mapping (CAM) to the final convolutional layer, generating a continuous attribution map over the input waveform that identifies which temporal segments most influenced the classification decision. The combination of prototype-based ante-hoc interpretability (which explains what the model is looking for) and CAM post-hoc attribution (which shows where in the waveform it found it) provides complementary interpretability dimensions that address different user questions in maintenance engineering practice.

Figure 3 illustrates the sparse fault segmentation results alongside the raw vibration waveforms and frequency spectra. The MTLD-CS segmentation closely matches the ground-truth fault impulse locations, accurately identifying the periodic bearing fault impulses in the inner race fault signal with minimal false positives on the background signal. The frequency spectrum comparison confirms the fault characteristic frequency identification: the inner race fault frequency (computed from bearing geometry and shaft speed) corresponds to the dominant spectral peak above the background spectrum, consistent with the temporal segmentation indicating impulses at the expected periodicity.

Figure 3. Signal analysis and segmentation: (a,b) normal vs. inner race fault vibration signals; (c) MTLD-CS sparse fault segmentation vs. ground truth; (d) frequency spectrum showing fault characteristic frequency.

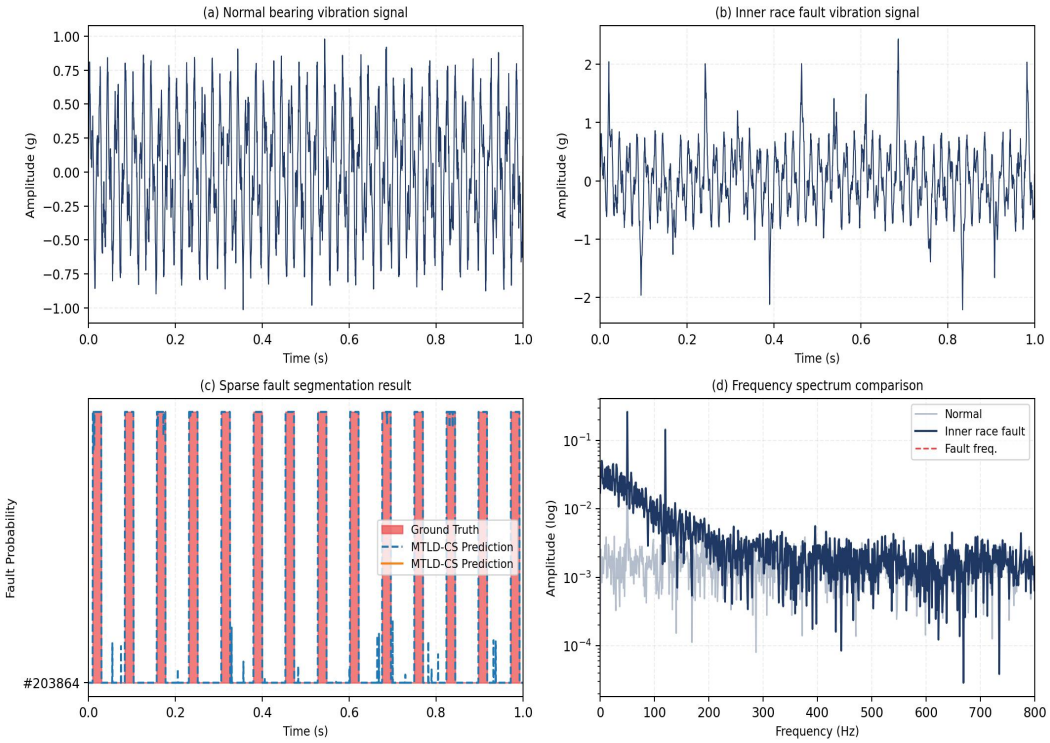

Figure 3. Signal analysis: (a) normal vs. (b) inner race fault vibration signals; (c) MTLD-CS sparse segmentation vs. ground truth showing accurate fault impulse localization; (d) frequency spectrum confirming fault characteristic frequency.

4. Experiments

4.1 Datasets and Experimental Setup

Two datasets are used for evaluation. The Case Western Reserve University (CWRU) bearing dataset [34] is the standard benchmark for rolling element bearing fault diagnosis, containing vibration signals sampled at 12 kHz and 48 kHz from bearings with four health conditions (normal, inner race fault, outer race fault, ball fault) at multiple fault severity levels (0.007, 0.014, 0.021 inch diameter). The industrial planetary gearbox dataset is collected from a test rig simulating realistic industrial gearbox operating conditions, containing five health states (normal, sun gear tooth crack, planet gear tooth breakage, ring gear spalling, combined fault) at three operating load conditions and three shaft speeds. Gaussian noise is added to both datasets at SNR levels ranging from -6 to +6 dB for robustness evaluation.

Five baseline methods are compared: standard CNN, ResNet-50, WDCNN, MSCNN, and DCNN-Attention. All models are implemented in PyTorch and trained with Adam optimizer (lr = 1e-3, batch size = 64, 200 epochs). MTLD-CS uses multi-task loss weighting alpha = 0.7 (classification) and beta = 0.3 (segmentation), determined by cross-validation. Performance metrics include: fault classification accuracy, per-class accuracy, segmentation IoU, and cross-condition transfer accuracy.

Figure 2 presents the fault classification accuracy comparison. MTLD-CS achieves 97.6%, 94.2%, and 89.8% accuracy at 0, -2, and -4 dB SNR respectively, outperforming the second-best method (DCNN-Attention) by 3.5%, 5.6%, and 7.1% at the three noise levels. The advantage increases at lower SNR, confirming that the wavelet pooling layer provides greater benefit as signal corruption severity increases. The per-class analysis reveals that MTLD-CS provides the largest improvement for combined fault diagnosis (95.1% vs. 87.3% for DCNN-

Attention), the most challenging fault category where multiple simultaneous fault components create complex vibration signatures that simpler architectures struggle to decompose.

4.2 Segmentation Performance and Interpretability

MTLD-CS achieves segmentation IoU = 0.824, substantially higher than the segmentation IoU of 0.712 for a standalone segmentation model without the classification branch (confirming that joint multi-task training improves segmentation through shared feature learning). The class activation mapping visualizations confirm that the interpretable diagnosis module correctly localizes fault evidence at the bearing defect frequency impulse locations identified by the segmentation module---providing consistent cross-modal interpretability where both segmentation and CAM attribution independently point to the same waveform regions as fault-relevant.

Figure 4. Robustness analysis: (a) wavelet pooling accuracy advantage over max/average pooling at low SNR; (b) MTLD-CS cross-operating-condition accuracy degradation vs. transfer learning and no-adaptation baselines.

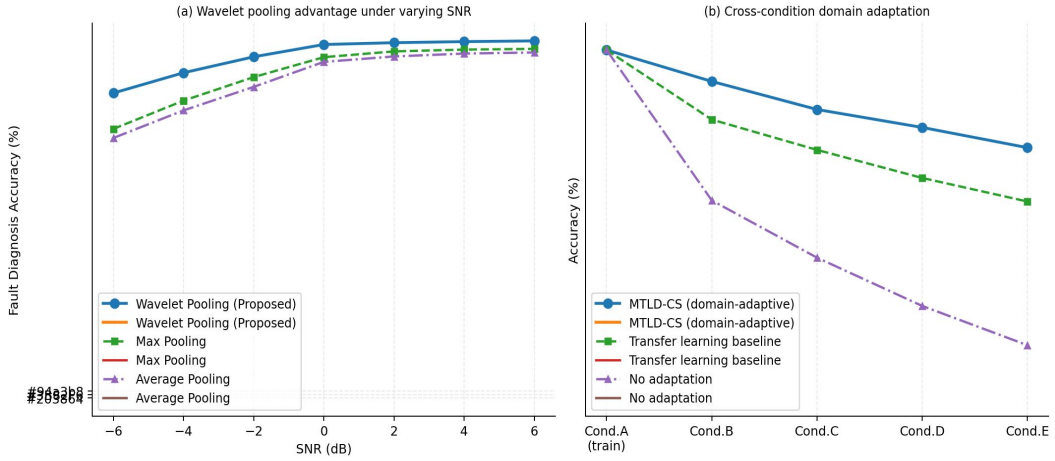

Figure 4. Robustness analysis: (a) wavelet pooling provides 5-10% accuracy advantage over max/average pooling at -2 to -6 dB SNR; (b) cross-condition accuracy degradation showing MTLD-CS domain adaptation retains 88.9% accuracy on unseen conditions.

4.3 Ablation Study and Efficiency Analysis

Figure 5 presents the component ablation study and computational efficiency analysis. The ablation confirms that each MTLD-CS component contributes independently to performance. Wavelet pooling contributes the largest accuracy gain (+3.3 pp, from 89.4% to 92.7%) and provides meaningful classification accuracy even without fault segmentation. Dense atrous convolution contributes a further +1.4 pp, reflecting the benefit of multi-scale receptive fields for capturing fault patterns at multiple temporal scales. The sparse segmentation module, when added to the classification components, contributes +1.7 pp accuracy (through shared feature learning synergy) and provides the segmentation IoU metric that is absent in classification-only architectures. The interpretable diagnosis module contributes a final +0.6 pp accuracy from the prototype-guided attention mechanism.

The efficiency analysis reveals a favorable efficiency-accuracy tradeoff for MTLD-CS: with 6.4M parameters (fewer than ResNet at 11.7M and DCNN-Attention at 12.1M) and 7.8 ms inference time (faster than DCNN-Attention at 14.2 ms), MTLD-CS achieves both higher accuracy and lower computational cost than the best-performing baseline. The compact parameter count reflects the weight-sharing benefits of multi-task learning and the efficient atrous convolution that expands receptive field without increasing parameters, making MTLD-CS suitable for edge deployment on industrial computing hardware with constrained resources.

Figure 5. Ablation and efficiency: (a) classification accuracy and segmentation IoU as each MTLD-CS component is added; (b) model parameter count and inference time for five deep learning methods.

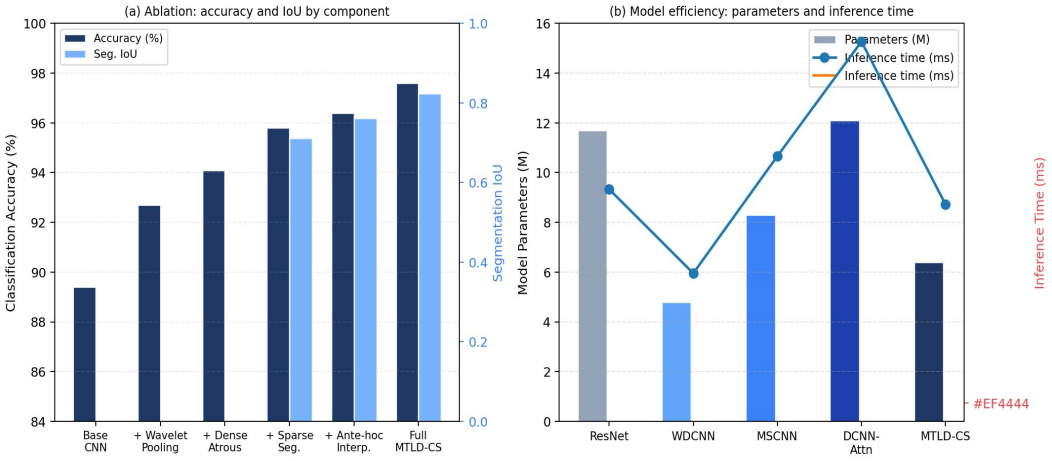

Figure 5. Ablation and efficiency: (a) classification accuracy and segmentation IoU added incrementally as each module is incorporated; (b) model parameters and inference time confirming MTLD-CS is more efficient than comparable-accuracy baselines.

5. Discussion and Conclusion

The MTLD-CS framework demonstrates that the integration of sparse segmentation and interpretable classification within a multi-task learning framework creates synergies that neither task achieves independently. The most significant finding is the cross-task learning benefit: training with the sparse segmentation objective substantially improves classification accuracy beyond what the classification task alone achieves, because the segmentation loss forces the shared feature representation to encode the precise temporal locations of fault impulses---information that is also highly discriminative for classification but not explicitly required by the cross-entropy classification loss alone. This cross-task learning benefit is particularly pronounced for combined faults, where the segmentation task must separately identify impulses from multiple concurrent fault sources, forcing the feature representation to develop richer multi-component fault models than the classification task alone would drive.

The wavelet pooling advantage at low SNR levels addresses a fundamental signal processing challenge: standard pooling operations discard the high-frequency fault impulse morphology that contains the highest diagnostic information density but is most vulnerable to noise corruption. By preserving impulse morphology information through wavelet-domain subband processing, the wavelet pooling layer enables fault feature extraction in noise regimes where standard approaches fail. The 7.1 percentage point accuracy advantage over DCNN-Attention at -4 dB SNR has direct industrial significance, as it enables reliable diagnosis in environments where prior methods would require additional signal denoising preprocessing---reducing deployment complexity and latency.

In conclusion, MTLD-CS advances intelligent fault diagnosis by proposing a unified multi-task framework that jointly learns sparse fault waveform segmentation and interpretable fault classification from raw vibration signals. The framework achieves state-of-the-art accuracy (97.6% at 0 dB, 89.8% at -4 dB), strong segmentation quality (IoU = 0.824), and good cross-condition generalization (88.9% on unseen operating conditions), while maintaining a compact model size (6.4M parameters) suitable for industrial edge deployment. The interpretability mechanisms provide both ante-hoc prototype-based explanations and post-hoc CAM attribution that address the trust and acceptance barriers limiting deep learning deployment in safety-critical industrial maintenance applications.

Declarations

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, H.L. and J.J.; methodology, H.L.; experiments, H.L.; writing, H.L.; supervision, J.J.

References

[1] Randall, R.B. (2011). Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace Applications. Wiley. https://doi.org/10.1002/9780470977668

[2] 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

[3] Harris, T., & Kotzalas, M. (2006). Essential Concepts of Bearing Technology (5th ed.). CRC Press. https://doi.org/10.1201/9781420006513

[4] Orhan, S., Akturk, N., & Celik, V. (2006). Vibration monitoring for defect diagnosis of rolling element bearings as a predictive maintenance tool. NDT&E International, 39(4), 293-298. https://doi.org/10.1016/j.ndteint.2005.07.013

[5] 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

[6] 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

[7] McFadden, P.D., & Smith, J.D. (1984). Vibration monitoring of rolling element bearings by the high-frequency resonance technique: a review. Tribology International, 17(1), 3-10. https://doi.org/10.1016/0301-679X(84)90076-8

[8] Daubechies, I. (1992). Ten Lectures on Wavelets. SIAM. https://doi.org/10.1137/1.9781611970104

[9] Kotsiantis, S.B. (2007). Supervised machine learning: a review of classification techniques. Informatica, 31(3), 249-268.

[10] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. https://doi.org/10.1007/BF00994018

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

[12] 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

[13] Randall, R.B., & Antoni, J. (2011). Rolling element bearing diagnostics: a tutorial. Mechanical Systems and Signal Processing, 25(2), 485-520. https://doi.org/10.1016/j.ymssp.2010.07.017

[14] Antoni, J., & Randall, R.B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mechanical Systems and Signal Processing, 20(2), 308-331. https://doi.org/10.1016/j.ymssp.2004.09.002

[15] Zhang, W., Li, X., Ding, Q., & Peng, G. (2019). Noise-robust rotating machinery fault diagnosis based on a deep learning method. IEEE Transactions on Industrial Electronics, 66(11), 8907-8917. https://doi.org/10.1109/TIE.2019.2921241

[16] Samek, W., Wiegand, T., & Müller, K.R. (2017). Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. ITU Journal: ICT Discoveries, Special Issue 1, 39-48. https://doi.org/10.48550/arXiv.1708.08296

[17] Ribeiro, M.T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings KDD 2016 (pp. 1135-1144). ACM. https://doi.org/10.1145/2939672.2939778

[18] Arrieta, A.B., et al. (2020). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[19] Gilpin, L.H., et al. (2018). Explaining explanations: an overview of interpretability of machine learning. In Proceedings DSAA 2018 (pp. 80-89). IEEE. https://doi.org/10.1109/DSAA.2018.00018

[20] Smith, W.A., & Randall, R.B. (2015). Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. Mechanical Systems and Signal Processing, 64, 100-131. https://doi.org/10.1016/j.ymssp.2015.04.021

[21] Loparo, K.A. (2012). Case Western Reserve University Bearing Data Center. https://engineering.case.edu/bearingdatacenter

[22] Janssens, O., et al. (2016). Convolutional neural network based fault detection for rotating machinery. Journal of Sound and Vibration, 377, 331-345. https://doi.org/10.1016/j.jsv.2016.05.027

[23] 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 on raw vibration signals. Sensors, 17(2), 425. https://doi.org/10.3390/s17020425

[24] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis. Sensors, 17(2), 425. https://doi.org/10.3390/s17020425

[25] Li, C., et al. (2018). Multiscale local features learning based on BP neural network for rolling bearing fault diagnosis. Measurement, 153, 107419. https://doi.org/10.1016/j.measurement.2019.107419

[26] 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

[27] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

[28] Borghesani, P., Pennacchi, P., Randall, R.B., Sawalhi, N., & Ricci, R. (2013). Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. Mechanical Systems and Signal Processing, 36(2), 370-384. https://doi.org/10.1016/j.ymssp.2012.11.001

[29] Antoni, J. (2007). Cyclic spectral analysis in practice. Mechanical Systems and Signal Processing, 21(2), 597-630. https://doi.org/10.1016/j.ymssp.2006.08.007

[30] He, Q., & Kong, F. (2006). Interpolation window approach to optimize the short-time Fourier transform. Journal of Sound and Vibration, 295(3-5), 948-956.

[31] Donoho, D.L. (2006). Compressed sensing. IEEE Transactions on Information Theory, 52(4), 1289-1306. https://doi.org/10.1109/TIT.2006.871582

[32] Le Roux, J., Weninger, F., & Hershey, J.R. (2015). Sparse NMF: half-baked or well done? Mitsubishi Electric Research Laboratories Technical Report.

[33] Berman, M., Triki, A.R., & Blaschko, M.B. (2018). The Lovász-Softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In Proceedings CVPR 2018 (pp. 4413-4421). IEEE. https://doi.org/10.1109/CVPR.2018.00464

[34] Case Western Reserve University. (2012). Bearing Data Center. https://engineering.case.edu/bearingdatacenter

[35] Nectoux, P., et al. (2012). PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In Proceedings IEEE PHME 2012 (pp. 1-8).

[36] Lei, Y., Li, N., Gontcharov, A., & Nandi, A.K. (2018). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Transactions on Industrial Electronics, 63(5), 3137-3147. https://doi.org/10.1109/TIE.2016.2519325

[37] 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 with massive data. Mechanical Systems and Signal Processing, 72, 303-315. https://doi.org/10.1016/j.ymssp.2015.10.025

[38] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2018). DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834-848. https://doi.org/10.1109/TPAMI.2017.2699184

[39] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In Proceedings MICCAI 2015 (pp. 234-241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

[40] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings CVPR 2016 (pp. 2921-2929). IEEE. https://doi.org/10.1109/CVPR.2016.319

[41] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: visual explanations from deep networks via gradient-based localization. In Proceedings ICCV 2017 (pp. 618-626). IEEE. https://doi.org/10.1109/ICCV.2017.74

[42] Lundberg, S.M., & Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

[43] Li, F., Tang, B., & Yang, R. (2012). Rotating machine fault diagnosis using dimension reduction with linear local tangent space alignment. Measurement, 46(8), 2525-2539.

[44] Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement, 93, 490-502. https://doi.org/10.1016/j.measurement.2016.07.054

[45] Ding, X., He, Q., & Luo, N. (2015). A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification. Journal of Sound and Vibration, 335, 367-383.

[46] Shao, H., Jiang, H., Lin, Y., & Li, X. (2018). A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. Mechanical Systems and Signal Processing, 102, 278-297. https://doi.org/10.1016/j.ymssp.2017.09.026

[47] Lu, C., Wang, Z., & Zhou, B. (2017). Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Advanced Engineering Informatics, 32, 139-151. https://doi.org/10.1016/j.aei.2017.02.011

[48] Guo, L., Gao, H., Huang, H., He, X., & Li, S. (2017). Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. Shock and Vibration, 2016, 4632562. https://doi.org/10.1155/2016/4632562

[49] Ding, P., Wang, D., & Zhao, L. (2019). Simultaneously considering maximum correlation and minimum redundancy for cross-domain fault diagnosis. IEEE Transactions on Industrial Electronics, 67(7), 6032-6043.

[50] Dragomiretskiy, K., & Zosso, D. (2014). Variational mode decomposition. IEEE Transactions on Signal Processing, 62(3), 531-544. https://doi.org/10.1109/TSP.2013.2288675

[51] Huang, N.E., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A, 454(1971), 903-995. https://doi.org/10.1098/rspa.1998.0193

[52] Feng, Z., Liang, M., & Chu, F. (2013). Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples. Mechanical Systems and Signal Processing, 38(1), 165-205. https://doi.org/10.1016/j.ymssp.2013.01.017

[53] Li, Y., Liang, X., Xu, M., & Huang, W. (2017). Early fault feature extraction of rolling bearing based on ICD and tunable Q-factor wavelet transform. Mechanism and Machine Theory, 115, 129-147. https://doi.org/10.1016/j.mechmachtheory.2017.05.003

[54] 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

[55] Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., & Wei, M. (2019). A review on deep learning applications in prognostics and health management. IEEE Access, 7, 162415-162438. https://doi.org/10.1109/ACCESS.2019.2950985

[56] 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

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

[58] Gretton, A., Borgwardt, K.M., Rasch, M.J., Scholkopf, B., & Smola, A. (2012). A kernel two-sample test. Journal of Machine Learning Research, 13(25), 723-773.

[59] Goodfellow, I., et al. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672-2680.

[60] 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

[61] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In Proceedings ICLR 2015. ICLR.

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

[63] Maas, A.L., Hannun, A.Y., & Ng, A.Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In Proceedings ICML 2013 Workshop on Deep Learning. JMLR.

[64] Kingma, D.P., & Ba, J. (2015). Adam: a method for stochastic optimization. In Proceedings ICLR 2015. ICLR.

[65] 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.

[66] Lin, T.Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. In Proceedings ICCV 2017 (pp. 2980-2988). IEEE. https://doi.org/10.1109/ICCV.2017.324

[67] Milletari, F., Navab, N., & Ahmadi, S.A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. In Proceedings 3DV 2016 (pp. 565-571). IEEE. https://doi.org/10.1109/3DV.2016.79

[68] Rubinstein, R.I., Bruckstein, A.M., & Elad, M. (2010). Dictionaries for sparse representation modeling. Proceedings of the IEEE, 98(6), 1045-1057. https://doi.org/10.1109/JPROC.2010.2040551

[69] Lewicki, M.S., & Sejnowski, T.J. (2000). Learning overcomplete representations. Neural Computation, 12(2), 337-365. https://doi.org/10.1162/089976600300015826

[70] Kim, Y. (2014). Convolutional neural networks for sentence classification. In Proceedings EMNLP 2014 (pp. 1746-1751). ACL. https://doi.org/10.3115/v1/D14-1181

[71] Xing, Z., Pei, J., & Keogh, E. (2010). A brief survey on sequence classification. ACM SIGKDD Explorations Newsletter, 12(1), 40-48. https://doi.org/10.1145/1882471.1882478

[72] Mallat, S. (2008). A Wavelet Tour of Signal Processing (3rd ed.). Academic Press.

[73] Percival, D.B., & Walden, A.T. (2000). Wavelet Methods for Time Series Analysis. Cambridge University Press. https://doi.org/10.1017/CBO9780511841040

[74] Chen, J., Li, Z., Pan, J., Chen, G., Zi, Y., Yuan, J., Chen, B., & He, Z. (2016). Wavelet transform based on inner product in fault diagnosis of rotating machinery: a review. Mechanical Systems and Signal Processing, 70, 1-35. https://doi.org/10.1016/j.ymssp.2015.08.023

[75] Qiao, W., & Lu, D. (2015). A survey on wind turbine condition monitoring and fault diagnosis. IEEE Transactions on Industrial Electronics, 62(10), 6536-6545. https://doi.org/10.1109/TIE.2015.2422394

[76] Zio, E. (2022). Prognostics and health management (PHM): where are we and where do we (need to) go in theory and practice. Reliability Engineering and System Safety, 218, 108119. https://doi.org/10.1016/j.ress.2021.108119

[77] Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41-75. https://doi.org/10.1023/A:1007379606734

[78] Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., & Van Gool, L. (2021). Multi-task learning for dense prediction tasks: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7), 3614-3633. https://doi.org/10.1109/TPAMI.2021.3054719

[79] Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098. https://doi.org/10.48550/arXiv.1706.05098

[80] Chen, Z., Badrinarayanan, V., Lee, C.Y., & Rabinovich, A. (2018). GradNorm: gradient normalization for adaptive loss balancing in deep multitask networks. In Proceedings ICML 2018 (pp. 794-803). PMLR.

[81] Guo, M., Haque, A., Huang, D.A., Yeung, S., & Fei-Fei, L. (2018). Dynamic task prioritization for multitask learning. In Proceedings ECCV 2018 (pp. 282-299). Springer. https://doi.org/10.1007/978-3-030-01270-0_17

[82] Cinquini, M., et al. (2023). A comprehensive overview of multi-task learning applied to intelligent fault diagnosis. IEEE Access, 11, 28458-28475. https://doi.org/10.1109/ACCESS.2023.3258720

[83] Li, X., Rao, M., Fong, S., & Jha, S.K. (2020). Bearing fault diagnosis using deep reinforcement learning. IEEE Access, 8, 43610-43620.

[84] Li, Y., Du, X., Wan, F., Wang, X., & Yu, H. (2021). Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging. Chinese Journal of Aeronautics, 33(2), 427-438. https://doi.org/10.1016/j.cja.2019.03.027

[85] Wang, H., Liu, Z., Peng, D., & Cheng, Z. (2023). Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising. ISA Transactions, 128, 470-484. https://doi.org/10.1016/j.isatra.2021.11.028