A Lightweight Neural Network for End-to-End Bearing Fault Diagnosis in Multi-Sensor Scenarios

Yichao Li[1], Yanfang Liu[1], and Xiangyang Xu[1]

[1] *School of Transportation Science and Engineering, Beihang University, Beijing, 100191, China*
leeyichao@buaa.edu.cn
liuyf@buaa.edu.cn
xxy@buaa.edu.cn

Abstract

Deep learning techniques have been widely applied in bearing fault diagnosis. However, their inherent reliance on historical offline training data and the large number of parameters pose considerable challenges in meeting the real-time requirements of online fault diagnosis applications, particularly in Industrial Internet of Things (IIoT) and edge computing environments. To address these challenges, this paper introduces a lightweight temporal feature fusion network (LTFFNet) for processing multi-sensor signals to enable end-to-end bearing fault diagnosis. Instead of following the prevalent approach of converting one-dimensional vibration signals into two-dimensional images for feature extraction and classification, we designed the architecture directly from the perspective of temporal signals. Besides, the incorporation of the Squeeze-and-Excitation (SE) module allows the network to adaptively recalibrate channel-wise feature responses. We assessed the accuracy and real-time performance of the developed network on an embedded platform using the CWRU bearing dataset. The results demonstrate high diagnostic capability and low computational time, indicating its effectiveness and suitability for real-time multi-sensor bearing fault diagnosis in industrial settings.

1. Introduction

Bearing fault diagnosis is a critical aspect of predictive maintenance in various industrial applications, playing a pivotal role in ensuring operational safety, minimizing downtime, and reducing maintenance costs (Chennana et al., 2025). Traditional approaches to fault diagnosis often rely on signal processing techniques coupled with expert knowledge, which can be labor-intensive and may lack robustness in complex real-world scenarios (Rai & Upadhyay, 2016). In recent years, deep learning techniques have emerged as a powerful tool for automated fault diagnosis due to their ability to automatically learn intricate features from raw sensor data, demonstrating superior performance compared to conventional methods (Lei, Li, Zhang, Wu, & Yu, 2025). These methods, leveraging architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown significant promise in identifying subtle fault patterns from vibration, acoustic, or current signals (Saeed et al., 2025).

Despite their success, existing deep learning models for bearing fault diagnosis face considerable challenges, particularly in the context of online and real-time applications within Industrial Internet of Things (IIoT) and edge computing environments (Siddique, Saleem, Umar, Kim, & Kim, 2025). A primary limitation stems from their inherent reliance on extensive historical offline training data, which may not always be readily available or representative of evolving operational conditions. Furthermore, many state-of-the-art deep learning architectures are characterized by a large number of parameters, leading to high computational complexity and memory consumption (Ma, Zhang, Zheng, & Sun, 2018). This poses significant hurdles for deployment on resource-constrained embedded platforms and edge devices, where real-time processing and rapid diagnostic responses are imperative. Consequently, the pursuit of lightweight yet accurate diagnostic models that can operate efficiently in such environments remains a critical research imperative.

To address these challenges, this paper introduces a novel lightweight temporal feature fusion network (LTFFNet) designed for end-to-end bearing fault diagnosis using multi-sensor signals. Unlike prevalent methodologies that involve converting one-dimensional vibration signals into two-dimensional (2D) images for subsequent feature extraction and classification, our proposed architecture is specifically engineered to process temporal signals directly. This time-series-centric design avoids potential information loss or artifacts introduced by 2D transformation and allows for a more

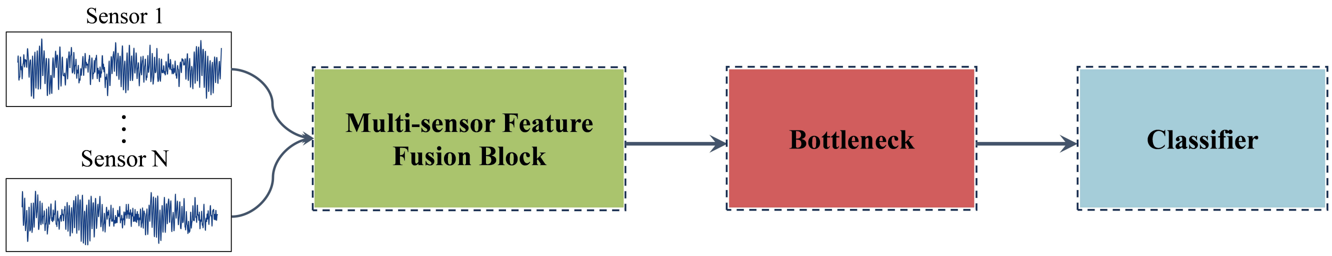

Figure 1. Overall architecture of LTFFNet

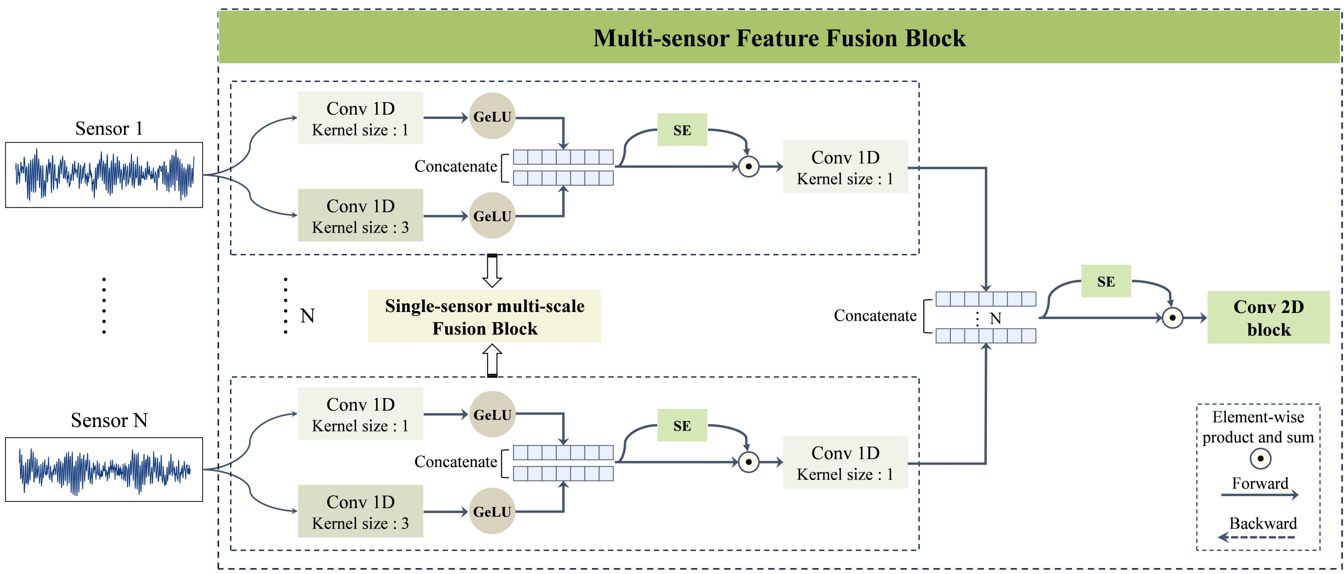

Figure 2. Multi-sensor feature fusion block

direct learning of temporal dependencies inherent in the sensor data. The incorporation of the Squeeze-and-Excitation (SE) module (Hu, Shen, Albanie, Sun, & Wu, 2020) enables the network to adaptively recalibrate channel-wise feature responses, thereby enhancing the representational power of the extracted features by emphasizing more informative channels and suppressing less relevant ones. This adaptive mechanism is particularly beneficial for effectively fusing information from diverse multi-sensor inputs.

The diagnostic accuracy and real-time performance of the developed LTFFNet were rigorously assessed on an embedded platform using the publicly available CWRU bearing dataset, a widely recognized benchmark in the field. Our experimental results unequivocally demonstrate that the proposed LTFFNet achieves high diagnostic capability while maintaining remarkably low computational time. These findings underscore the effectiveness and suitability of our lightweight network for practical, real-time multi-sensor bearing fault diagnosis applications in demanding IIoT and edge computing environments.

2. PROPOSED METHOD

This section details the architecture of the proposed Lightweight Temporal Feature Fusion Network (LTFFNet) designed for end-to-end bearing fault diagnosis using multi-sensor data. The overall architecture of the LTFFNet is depicted in Fig. 1, illustrating its modular design which facilitates efficient processing and learning from complex time-series signals. The LTFFNet is fundamentally composed of three main sequential modules: a Multi-sensor Feature Fusion Block, a Bottleneck module, and a Classifier.

2.1. Overall Architecture of LTFFNet

As shown in Fig. 1, the raw time-series signals from multiple sensors (Sensor 1 to Sensor N) serve as the input to the LTFFNet. These inputs are first processed by the multi-sensor feature fusion block. This block is designed to effectively integrate and extract relevant features from heterogeneous multi-sensor data, transforming the raw temporal signals into a unified, high-level feature representation. The output of the multi-sensor feature fusion block is then fed into the bottleneck module. The bottleneck module serves to compress the feature representation, reduce dimensionality, and enhance the network's efficiency by distilling the most critical information while mitigating overfitting. Finally, the condensed features from the bottleneck module are passed to the classifier, which is responsible for performing the final fault diagnosis by mapping the learned features to specific fault categories.

2.2. Multi-sensor Feature Fusion Block

The multi-sensor feature fusion block, a core component of the LTFFNet, is illustrated in detail in Fig. 2. This block is crucial for leveraging the complementary information available from different sensing modalities. It comprises two primary stages: a single-sensor multi-scale fusion stage and a cross-sensor fusion stage. For each individual sensor input (e.g., Sensor 1 to Sensor N), a single-sensor multi-scale fusion block is employed. Within each single-sensor block, the raw 1D temporal signal is processed by parallel convolutional pathways to capture features at different scales. Specifically, two distinct 1D convolution (Conv 1D) layers are used: one with a kernel size of 1 and another with a kernel size of 3. These convolutions are followed by a GeLU activation function, which introduces non-linearity. The outputs from these parallel pathways are then concatenated to form a richer, multi-scale feature representation for that specific sensor.

Crucially, Squeeze-and-Excitation (SE) modules are incorporated after the concatenation within each single-sensor multi-scale fusion block. The SE module, originally proposed by Hu et al. (Hu et al., 2020), plays a vital role in adaptively recalibrating channel-wise feature responses. This mechanism allows the network to automatically learn the importance of each feature channel, thereby emphasizing more informative channels and suppressing less relevant ones. This selective enhancement significantly boosts the representational power of the features extracted from individual sensors. Following the SE module, another 1D convolution with a kernel size of 1 further refines these fused features.

After individual multi-scale feature extraction and recalibration for each sensor, the outputs from all single-sensor multi-scale fusion blocks are concatenated to form a comprehensive multi-sensor feature map. This concatenated feature map then undergoes another SE module for global channel-wise recalibration across all fused sensor features. Finally, a 2D convolution block (Conv 2D block) processes this globally recalibrated multi-sensor feature map, producing the final output of the multi-sensor feature fusion block. This hierarchical fusion strategy, combining multi-scale feature extraction at the single-sensor level with adaptive channel recalibration and subsequent cross-sensor fusion, ensures that the LTFFNet can effectively capture and integrate critical diagnostic information from diverse multi-sensor inputs.

3. EXPERIMENT

3.1. CWRU Bearing Dataset

The performance of the proposed LTFFNet was validated using the widely recognized Case Western Reserve University (CWRU) bearing dataset. This dataset serves as a standard benchmark for rotating machinery fault diagnosis, collected from a test rig comprising an electric motor, a torque transducer/encoder, and a dynamometer, as illustrated in Fig. 3. For this study, vibration data sampled at 12 kHz from both the fan-end and drive-end bearing sensors were utilized to verify the effectiveness of the LTFFNet in multi-sensor scenarios.

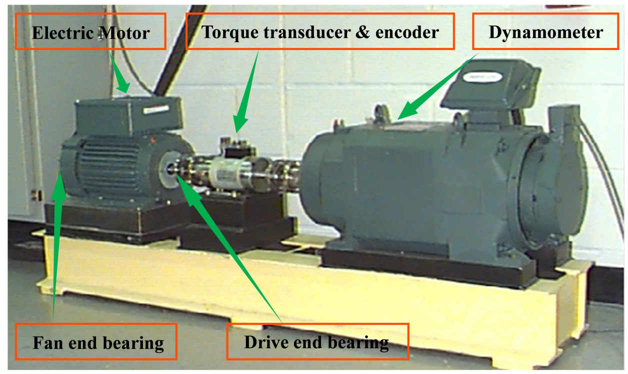

Figure 3. CWRU test rig

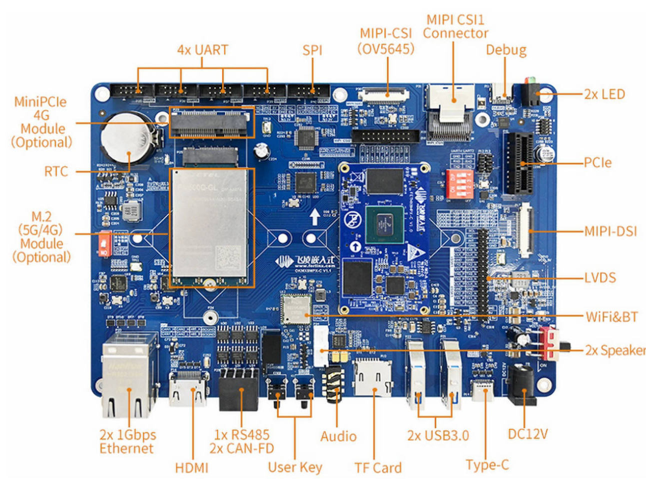

Figure 4. OKMX8MPQ-C platform

3.2. Embedded Platform

To assess the real-time performance and suitability of the LTFFNet for edge computing applications, experiments were conducted on an embedded platform. The chosen platform is the OKMX8MPQ-C development board, as shown in Fig. 4. This board is built around the NXP i.MX 8M Plus high-performance processor, which integrates a Neural Processing Unit (NPU) and an Image Signal Processor (ISP). The NPU

Table 1. Diagnostic accuracy and inference time of LTFFNet under different load conditions

Load (hp)	Accuracy	Avg. inference time / batch (ms)	std. dev. (ms)	Throughput (Samples/sec)
0	99.33%	292.37	12.76	109.45
1	98.40%	290.73	9.08	110.07
2	97.39%	293.15	14.45	109.16
3	99.16%	297.82	7.06	107.43

provides an AI computing capability of up to 2.3 TOPS (Tera Operations Per Second), making the OKMX8MPQ-C development board well-suited for lightweight edge computing demands. This powerful yet efficient hardware platform allows for a realistic evaluation of the LTFFNet's deployment feasibility in practical IIoT environments.

3.3. Results and Analysis

The efficacy of the proposed LTFFNet was comprehensively evaluated in terms of both diagnostic accuracy and computational efficiency, crucial metrics for deployment on edge devices. The model, with a total of approximately 2.05 Mb trainable parameters, was trained using a batch size of 32 and a learning rate of 0.001. The performance was assessed under four distinct motor load conditions (0, 1, 2, and 3 hp).

The quantitative results, presented in Table. 1, demonstrate the exceptional performance of the LTFFNet. The network achieved consistently high diagnostic accuracy across all tested operational loads, ranging from 97.39% to a peak of 99.33%. This robustness to varying load conditions is critical for real-world industrial applications where machinery often operates under dynamic states. In terms of computational performance, its efficiency is highlighted by a mean batch inference latency between 290.73 ms and 297.82 ms. This enables a stable processing throughput of approximately 107 to 110 samples per second, confirming the model's capacity to handle continuous data streams effectively. Furthermore, the low standard deviation of the inference time (ranging from 7.06 ms to 14.45 ms) signifies high temporal consistency, a critical attribute for predictable real-time systems. This performance, translating to a per-sample latency of approximately 9.09–9.31 ms, validates the proposed architecture as a practical and reliable solution for fault diagnosis in IIoT and edge computing environments.

To qualitatively assess the feature learning capability of the LTFFNet, we utilized the t-SNE technique to visualize the high-dimensional features extracted by the network immediately preceding the final classifier layer. Fig. 5 illustrates the feature distribution for the 0 hp load condition. The visualization reveals ten distinct and well-separated clusters, each corresponding to a specific health state of the bearing (one normal and nine fault categories). The high degree of intra-class compactness and inter-class separability evident in the plot indicates that the multi-sensor feature fusion block effectively learns highly discriminative representations from the raw temporal signals. This strong feature separability provides a clear explanation for the high diagnostic accuracy reported in Table. 1 and validates the architectural design of the network.

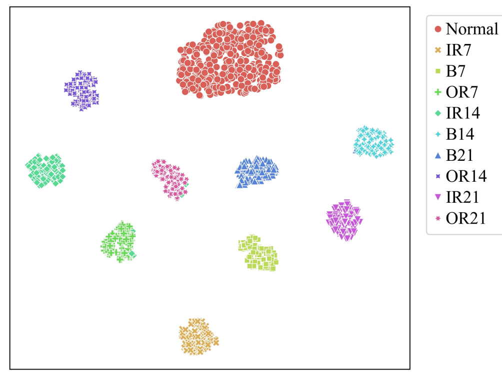

Figure 5. t-SNE visualization of learned features under 0 hp load condition

4. CONCLUSION

This paper addresses the critical challenge of deploying accurate and efficient fault diagnosis models on resource-constrained edge devices for real-time industrial applications. We proposed a novel Lightweight Temporal Feature Fusion Network (LTFFNet), an end-to-end architecture designed to directly process and integrate multi-sensor temporal signals for bearing fault diagnosis.

Experimental validation on the CWRU benchmark dataset demonstrated the outstanding performance of our proposed model. The LTFFNet achieved diagnostic accuracies consistently above 97% across various load conditions, showcasing its robustness and reliability. Furthermore, when deployed on the NXP i.MX 8M Plus-based embedded platform, the model exhibited remarkably low inference latency, affirming its feasibility for real-time implementation.

In conclusion, the LTFFNet provides an effective and practical solution for end-to-end bearing fault diagnosis in multi-sensor scenarios. Its lightweight design and high performance make it particularly well-suited for deployment in edge computing environments, paving the way for more intelligent and autonomous predictive maintenance systems.

ACKNOWLEDGEMENT

This work was supported by the Basic Research Program (DEDP2023007) of the Ministry of Industry and Information Technology, China.

BIOGRAPHIES

Yichao Li received the M.S. degree in control theory and control science from Xiamen University, Xiamen, China, in 2022. He is currently working toward the Ph.D. degree in automotive engineering. His research focuses on industrial artificial intelligence for mechanical systems, specializing in fault diagnosis and remaining useful life prediction.

Yanfang Liu received the Ph.D. degree in mechatronic engineering from the Beihang university, Beijing, China, in 2005. She is a professor of school of transportation science and technology, Beihang University, Beijing, China. Her research interests include intelligent control technology of transmission systems, prognostics and health management (PHM).

Xiangyang Xu received the Ph.D. degree in mechatronic engineering from the Harbin Institute of Technology, Harbin, China, in 1999. He is a professor of school of transportation science and technology, Beihang University, Beijing, China. Dr. Xu is the recipient of the First Prize of the National Science and Technology Progress Award as the first inventor in 2016.

REFERENCES

Chennana, A., Megherbi, A. C., Bessous, N., Sbaa, S., Teta, A., Belabbaci, E. O., … Agajie, T. F. (2025). Vibration signal analysis for rolling bearings faults diagnosis based on deep-shallow features fusion. *Scientific Reports*, *15*(1), 9270.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020, August). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, *42*(8), 2011–2023.

Lei, L., Li, W., Zhang, S., Wu, C., & Yu, H. (2025). Research progress on data-driven industrial fault diagnosis methods. *Sensors*, *25*(9).

Ma, N., Zhang, X., Zheng, H.-T., & Sun, J. (2018). Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Computer vision – eccv 2018: 15th european conference, munich, germany, september 8–14, 2018, proceedings, part xiv* (p. 122–138). Berlin, Heidelberg: Springer-Verlag.

Rai, A., & Upadhyay, S. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, *96*, 289-306.

Saeed, A., A. Khan, M., Akram, U., J. Obidallah, W., Jawed, S., & Ahmad, A. (2025). Deep learning based approaches for intelligent industrial machinery health management and fault diagnosis in resource-constrained environments. *Scientific Reports*, *15*(1), 1114.

Siddique, M. F., Saleem, F., Umar, M., Kim, C. H., & Kim, J.-M. (2025). A hybrid deep learning approach for bearing fault diagnosis using continuous wavelet transform and attention-enhanced spatiotemporal feature extraction. *Sensors*, *25*(9).