Deep Reinforcement Learning-Based Early Fault Warning for Rolling Bearings in Weak Data Scenarios

Mingyu Yue *

School of Mechanical Engineering, University of Leeds, Woodhouse Lane, Leeds, UK

* Corresponding Author Email: hlld8753@leeds.ac.uk

Abstract. Rolling bearings are vital components in rotating machinery, and early detection of bearing faults is essential to avoid unplanned downtime, costly repairs, and safety hazards. In industrial environments, early fault signatures are often weak, noisy, and rare-creating a “weak-data” scenario marked by severe class imbalance, high annotation cost, and distribution shifts that undermine conventional deep learning methods. Deep reinforcement learning (DRL) offers a promising alternative by framing fault warning as a Markov decision process and leveraging sequential interaction and reward signals to learn from limited and imbalanced data. This paper presents a systematic the paper of DRL-based approaches for early bearing-fault warning, categorizing and analyzing key enabling techniques: reward shaping and balanced experience replay, data augmentation and generative synthesis, integration of deep feature extractors (CNNs/LSTMs), meta-reinforcement learning for few-shot adaptation, transfer and domain adaptation, exploration strategies, and partially observable modelling. The paper further discusses practical challenges—including reward design, sample efficiency, training stability, interpretability, and computational constraints for edge deployment—and outlines promising directions such as generative-model augmentation, attention and graph-based architecture, multi-sensor fusion, and lightweight distributed DRL. This paper aims to guide the development of data-efficient, robust DRL solutions for industrial predictive maintenance.

Keywords: Deep Reinforcement Learning, Fault Diagnosis, Rolling Bearing, Weak Data.

1. Introduction

Rolling bearings are among the most critical components in rotating machinery. Their condition directly affects overall equipment performance, production continuity, and operational safety. Failure can easily lead to unplanned shutdowns, production halts, and expensive repairs—resulting in substantial economic losses and even safety risks [1, 2]. Therefore, achieving early fault warning in bearings is essential for advancing predictive maintenance and ensuring both safety and cost-effective operation [3].

In real industrial environments, early bearing fault signals—such as vibrations, acoustics, or current—are often very weak and exhibit low signal-to-noise ratios. They are easily masked by noise and varying operating conditions, making it difficult to detect using traditional threshold- or spectrum-based methods [4]. More importantly, acquiring sufficient and well-labelled fault data in practical settings is challenging, especially since data from early degradation stages are incredibly scarce. The paper refers to this type of constrained data environment as a "weak-data scenario." Its core challenges include scarce fault samples, high labelling costs, class imbalance, noise interference, and poor cross-condition adaptability [5]. These limitations significantly restrict the effectiveness of conventional deep learning models like CNNs and DNNs, often leading to overfitting and limited generalization [6].

Deep Reinforcement Learning (DRL) offers a new perspective for early fault warning under weak-data conditions. Based on agent-environment interaction, DRL uses reward signals to guide policy optimization and sequential decision-making [4, 5]. It can provide weak supervision through reward shaping, generate synthetic data via simulation or generative models, and enhance the ability to detect subtle fault patterns through intrinsic motivation mechanisms. This helps alleviate issues related to data scarcity and labelling difficulty.

This paper presents a systematic overview of deep reinforcement learning (DRL)-based methods for early fault warning in rolling bearings under weak-data scenarios. The paper focuses on key enabling techniques, including reward design, sample augmentation, meta-learning, transfer and domain adaptation, exploration strategies, and partially observable modelling. In addition, it discusses practical implementation challenges and suggests promising future research directions. The remainder of the paper is structured as follows: Section 2 introduces the characteristics of bearing faults and the fundamentals of DRL; Section 3 shows and categorises state-of-the-art DRL methods for fault diagnosis and warning; Section 4 analyses typical applications and implementation challenges.

2. Background and Challenges of Bearing Fault Warning Under Weak-Data Conditions

In real industrial environments, standard samples vastly outnumber fault samples, leading to highly imbalanced data distribution. This is especially true during the initial stages of equipment degradation, where valid fault examples are particularly scarce [5,7]. Acquiring samples that cover multiple fault types, varying severity levels, and complete life cycles remains a significant challenge [8]. Because destructive tests are usually avoided in actual operations, naturally degraded fault data are even more complex to obtain. What's more, fault annotation heavily depends on domain experts—not only is it costly and time-consuming, but consistency in labelling quality is also hard to guarantee.

Environmental noise, variable operating conditions, and non-ideal sensor characteristics further mask already subtle fault features [4]. Although public datasets such as those from CWRU (Case Western Reserve University) and Paderborn University have supported much of the current research, their data mainly come from accelerated tests under controlled conditions. These datasets often lack real-world variability—such as fluctuating loads, compound fault patterns, and natural degradation processes. This frequently causes trained models to encounter data distribution shifts and show poor generalisation when deployed in actual field applications.

Traditional signal processing methods, including threshold detection and spectral analysis, rely heavily on expert knowledge and tend to perform poorly under low signal-to-noise ratios and non-stationary conditions. While supervised deep learning models like CNNs and DNNs offer strong feature extraction capabilities, their performance depends extensively on large amounts of high-quality labelled data [9]. In weak-data environments characterized by scarce samples, labelling noise, and class imbalance, these models are prone to overfitting and exhibit limited generalisation.

Therefore, solely pursuing higher accuracy metrics without considering sample efficiency, noise robustness, and cross-condition adaptability will likely restrict the practical usefulness of such models.

3. The Potential and Advantages of DRL for Fault Early Warning

Fault early warning can be cast naturally as a Markov Decision Process (MDP), whose three core elements are the state—features extracted from vibration signals; the action—diagnostic decisions or graded alert levels; and the reward—typically shaped to reflect both accuracy and timeliness of diagnosis. It fits time-series decision problems particularly well. By learning through sequential interactions, an agent can optimize long-term outcomes and thus track evolving fault patterns over time; in practice, this sequential perspective often outperforms snapshot classification. What makes deep reinforcement learning (DRL) especially appealing here is its interaction-driven learning: feedback from reward signals guides policy improvement, thereby reducing reliance on large pre-labelled datasets. A carefully designed reward can make all the difference. Indeed, even with very few fault samples, an agent may learn to separate normal from abnormal behaviour, provided the reward highlights subtle, early deviations [5]. It should be noted that class imbalance remains a practical hurdle; to mitigate it, methods such as Dual Experience Pool Deep Reinforcement Learning (DEPDRL) have been proposed [10]. These keep separate replay buffers for majority and minority

classes while adopting balanced sampling, and—empirically—produce marked gains in diagnostic performance under imbalanced conditions.

What really stands out with DRL is its ability to automatically pull-out meaningful fault features from raw or lightly processed vibration signals—using deep neural networks. That means no tedious manual feature engineering, which usually demands deep expert knowledge [4, 9, 11]. This is especially important for detecting faint early faults, often hidden within complex time-frequency patterns [12, 13].

In fact, traditional fault diagnosis often leans heavily on expert-designed features, while deep learning—and DRL especially—simplifies the process through automated feature learning and data-driven classification [11].

DRL also brings built-in solutions to data imbalance. Reward shaping, for example, can motivate the agent to pay more attention to rare faults by offering higher rewards for correct identification [9]. Balanced experience replay also helps improve learning from infrequent fault patterns. In real applications, industrial machinery like bearings often runs under changing conditions. Variable speeds and loads can alter signal behaviour [8]. DRL agents are capable of learning online and adapting continuously to new situations. That dramatically improves the generalisation of diagnostic models [1, 5]. Better still, DRL doesn't just classify known faults. It learns patterns of "anomalous" behaviour, enabling real early warning. By treating early faults as slight state deviations, the agent can detect subtle irregularities and trigger alerts before failure develops [10]. Practical applications—like sensor-based early warning in power pipelines—already show the usefulness of this approach [14].

4. Research Progress on DRL-Based Early Fault Warning Methods

Currently, the application of Deep Reinforcement Learning (DRL) in early diagnosis and warning of bearing faults—especially under "weak-data" conditions characterized by faint signals and scarce labels—is rapidly becoming a research hotspot in the field of industrial intelligent maintenance. In real industrial settings, early fault samples are scarce, and most equipment remains in a normal state for long periods, leading to severely imbalanced data distribution. It is precisely this reality that has driven the development of a range of improved DRL methods.

At the algorithmic level, many researchers have focused on enhancing experience replay mechanisms. For example, Kang proposed a dual experience replay structure, which cleverly uses two buffers to store state transition samples from normal and fault conditions, respectively. Balanced cross-sampling then ensures the model is exposed to both "normal" and "abnormal" patterns. At the same time, reward functions have been redesigned—assigning higher rewards for correctly identifying rare faults— guiding the agent to actively focus on those subtle yet critical early signs of failure.

But strategy optimization alone is not enough; feature extraction is also evolving. Wang integrated deep networks such as CNN or LSTM into the DRL framework as feature extractors, enabling end-to-end learning from raw vibration signals to fault decisions. The benefit is clear: this eliminates the heavy manual effort tied to traditional feature engineering and reduces information loss during multi-stage processing. If these methods can be thought of as "using more data to solve data-scarcity", the introduction of Meta-Learning marks a real step toward "rapid adaptation with few samples". Meta-Reinforcement Learning (Meta-RL) allows an agent to learn a general policy across multiple related tasks. When encountering a new fault, it quickly adapts with minimal samples [8,15]. In essence, it learns the ability to "generalize from few examples"—a mechanism highly suited to industrial settings where new fault modes constantly emerge.

Even more interestingly, some studies have moved beyond "classification" and reframed fault warning as an "anomaly detection" problem. Since fault samples are hard to obtain, these methods use abundant normal data to train a "normality model". Any state deviating from this typical pattern

is treated as an anomaly [10]. This approach is not only technically feasible, but also more practical to deploy—after all, collecting normal data is much easier.

Beyond that, some researchers have experimented with enhancing signal preprocessing. For instance, Wang combined stochastic resonance with DRL to amplify weak fault features masked initially by noise, thereby improving the signal-to-noise ratio. It's like equipping the DRL agent with a "magnifying glass" before decision-making, making identification significantly more effective.

From a broader perspective, current research on DRL-based early fault warning is advancing through multiple integrated pathways. Whether through sampling strategies, reward mechanisms, feature learning architectures, meta-learning, or enhanced signal processing techniques, the ultimate goal remains the same: to make intelligent maintenance systems reliable and sensitive—even in real-world environments where data is scarce, noise is high, and operating conditions are complex.

5. Discussion

Although Deep Reinforcement Learning (DRL) offers a promising approach to early fault warning in bearings—especially in "weak-data" scenarios where fault samples are scarce and signals are subtle—several real and stubborn challenges remain. For one, reward design—a classic issue in DRL—becomes especially tricky here. Since early faults often develop gradually and invisibly, it's challenging to capture their progression using clear and timely reward signals [16]. A poorly designed reward can easily mislead the agent, resulting in nothing better than a suboptimal policy.

What's more, DRL training still demands a substantial amount of interactive data. Sample efficiency hasn't yet reached a practical level, and the training process itself is often unstable with uncertain convergence—all of which hinder real-world deployment. Then there's the "black-box" nature of DRL models: their decision-making lacks transparency, making it hard to gain the trust of engineers in high-stakes industrial settings. Add to that the high computational costs of training and inference, and it becomes clear why many current methods struggle to run on resource-limited edge devices. Yet these challenges are also driving new ideas. Several research directions look particularly promising. One approach is to combine transfer learning and domain adaptation, creating a "pre-train and fine-tune" framework that allows features and policies learned in one context to be transferred to new scenarios with scarce data [17-19].

On the architecture side, incorporating attention mechanisms could help DRL focus on key features. Graph neural networks could model structural relationships between components, and causal models might uncover hidden mechanisms behind faults. These new structures could improve both performance and interpretability [18,19]. Some researchers suggest using generative models—like GANs or diffusion models—to create realistic fault samples and ease data scarcity [20]. Meanwhile, with the rise of edge AI, lightweight and low-latency distributed DRL frameworks are gaining attention. The goal is to push diagnostic capabilities closer to the devices themselves. Going even further, integrating multi-sensor data—vibration, temperature, acoustic emission, and more—would give DRL agents a richer, more robust understanding of the environment [10,21]. That could significantly improve the accuracy and reliability of early warnings.

This paper has systematically examined the application of Deep Reinforcement Learning (DRL) for early fault warning in rolling bearings, explicitly addressing the critical challenge of "weak-data scenarios." These scenarios, defined by scarce fault samples, severe class imbalance, high noise, and labelling difficulties, significantly hinder the performance of traditional data-driven methods. The paper established that DRL, with its unique paradigm of learning through environmental interaction and reward signals, provides a powerful and flexible framework for sequential decision-making under these constraints. By formulating fault detection as a Markov Decision Process (MDP), DRL agents can learn optimal policies for identifying subtle, early-stage faults from raw sensory data with minimal supervision.

The core of the paper categorized and analysed key technical advancements designed to overcome data scarcity. These include innovative reward shaping to guide agent focus, enhanced experience

replay mechanisms (e.g., DEPDRL) to handle data imbalance, and the integration of deep feature extractors (CNNs, LSTMs) for end-to-end learning. Furthermore, the paper highlighted the pivotal role of meta-learning (Meta-RL) for few-shot adaptation to new faults and the reframing of the problem as anomaly detection, which leverages abundant normal data. The combination of DRL with advanced signal processing techniques, such as stochastic resonance, was also discussed to amplify weak fault features before decision-making.

Despite its promise, the practical deployment of DRL faces persistent challenges. These include the inherent difficulty in designing reward functions for gradual fault progression, the relatively low sample efficiency and training instability of DRL algorithms, a lack of model interpretability ("black box" nature), and high computational costs. To address these limitations, future research directions are outlined, focusing on the integration of transfer and domain adaptation, the incorporation of attention mechanisms and graph neural networks for improved performance and explainability, the use of generative models (GANs, diffusion models) for synthetic data augmentation, and the development of lightweight frameworks for edge computing. The fusion of multi-sensor data is identified as a key path toward building more robust and reliable systems. In conclusion, this paper underscores the significant potential of DRL to advance the state-of-the-art in predictive maintenance. It serves as a foundational reference for developing intelligent, data-efficient early warning systems for industrial machinery. This paper has systematically examined the application of Deep Reinforcement Learning (DRL) for early fault warning in rolling bearings, explicitly addressing the critical challenge of "weak-data scenarios." These scenarios, defined by scarce fault samples, severe class imbalance, high noise, and labelling difficulties, significantly hinder the performance of traditional data-driven methods. The paper established that DRL, with its unique paradigm of learning through environmental interaction and reward signals, provides a powerful and flexible framework for sequential decision-making under these constraints. By formulating fault detection as a Markov Decision Process (MDP), DRL agents can learn optimal policies for identifying subtle, early-stage faults from raw sensory data with minimal supervision.

6. Conclusion

In conclusion, the paper categorises and analyses key technical advancements designed to overcome data scarcity. These include innovative reward shaping to guide agent focus, enhanced experience replay mechanisms (e.g., DEPDRL) to handle data imbalance, and the integration of deep feature extractors (CNNs, LSTMs) for end-to-end learning. Furthermore, the paper highlighted the pivotal role of meta-learning (Meta-RL) for few-shot adaptation to new faults and the reframing of the problem as anomaly detection, which leverages abundant normal data. The combination of DRL with advanced signal processing techniques, such as stochastic resonance, was also discussed as a means to amplify weak fault features before decision-making.

Despite its promise, the practical deployment of DRL faces persistent challenges. These include the inherent difficulty in designing reward functions for gradual fault progression, the relatively low sample efficiency and training instability of DRL algorithms, a lack of model interpretability ("black-box" nature), and high computational costs. To address these limitations, future research directions are outlined, focusing on the integration of transfer and domain adaptation, the incorporation of attention mechanisms and graph neural networks for improved performance and explainability, the use of generative models (GANs, diffusion models) for synthetic data augmentation, and the development of lightweight frameworks for edge computing. The fusion of multi-sensor data is identified as a key path toward building more robust and reliable systems. In conclusion, this paper underscores the significant potential of DRL to advance the state-of-the-art in predictive maintenance. It serves as a foundational reference for developing intelligent, data-efficient early warning systems for industrial machinery. This paper has systematically examined the application of Deep Reinforcement Learning (DRL) for early fault warning in rolling bearings, explicitly addressing the critical challenge of "weak-data scenarios." These scenarios, defined by scarce fault samples, severe

class imbalance, high noise, and labelling difficulties, significantly hinder the performance of traditional data-driven methods. The paper established that DRL, with its unique paradigm of learning through environmental interaction and reward signals, provides a powerful and flexible framework for sequential decision-making under these constraints. By formulating fault detection as a Markov Decision Process (MDP), DRL agents can learn optimal policies for identifying subtle, early-stage faults from raw sensory data with minimal supervision.

The core of the paper categorized and analysed key technical advancements designed to overcome data scarcity. These include innovative reward shaping to guide agent focus, enhanced experience replay mechanisms (e.g., DEPDRL) to handle data imbalance, and the integration of deep feature extractors (CNNs, LSTMs) for end-to-end learning. Furthermore, the paper highlighted the pivotal role of meta-learning (Meta-RL) for few-shot adaptation to new faults and the reframing of the problem as anomaly detection, which leverages abundant normal data. The combination of DRL with advanced signal processing techniques, such as stochastic resonance, was also discussed as a means to amplify weak fault features before decision-making.

Despite its promise, the practical deployment of DRL faces persistent challenges. These include the inherent difficulty in designing reward functions for gradual fault progression, the relatively low sample efficiency and training instability of DRL algorithms, a lack of model interpretability ("black box" nature), and high computational costs. To address these limitations, future research directions are outlined, focusing on the integration of transfer and domain adaptation, the incorporation of attention mechanisms and graph neural networks for improved performance and explainability, the use of generative models (GANs, diffusion models) for synthetic data augmentation, and the development of lightweight frameworks for edge computing. The fusion of multi-sensor data is identified as a key path toward building more robust and reliable systems. In conclusion, this paper underscores the significant potential of DRL to advance the state-of-the-art in predictive maintenance. It serves as a foundational reference for developing intelligent, data-efficient early warning systems for industrial machinery.

References

[1] *Çakır, E., Dumond, P. (2025) A comparative analysis of reinforcement learning and conventional deep learning approaches for bearing fault diagnosis. arXiv preprint arXiv: 2506.19929.*

[2] *Wang, S., Li, J., Xu, X., Wu, R., Qiu, Y., Chen, X., Qiao, Z. (2025) Optimization of a coupled neuron model based on deep reinforcement learning and application of the model in bearing fault diagnosis. Sensors, 25 (12), 3654.*

[3] *Pham, M. T., Kim, J. M., Kim, C. H. (2020) Deep learning-based bearing fault diagnosis method for embedded systems. Sensors, 20 (23), 6886.*

[4] *Wang, R., Jiang, H., Zhu, K., Wang, Y., Liu, C. (2022) A deep feature enhanced reinforcement learning method for rolling bearing fault diagnosis. Advanced Engineering Informatics, 54, 101750.*

[5] *Li, Y., Wang, Y., Zhao, X., Chen, Z. (2024) A deep reinforcement learning-based intelligent fault diagnosis framework for rolling bearings under imbalanced datasets. Control Engineering Practice, 145, 105845.*

[6] *Ding, J., Li, X., Gudivada, V. N. (2017) Augmentation and evaluation of training data for deep learning. In Proceedings of the IEEE International Conference on Big Data, pp. 2603 – 2611. IEEE.*

[7] *Lin, L., Wang, B., Qi, J., Wang, D., Huang, N. (2019) Bearing fault diagnosis considering the effect of imbalance training sample. Entropy, 21 (4), 386.*

[8] *Wang, S., Wang, D., Kong, D., Wang, J., Li, W., Zhou, S. (2020) Few-shot rolling bearing fault diagnosis with metric-based meta learning. Sensors, 20 (22), 6437.*

[9] *Hoang, D. T., Kang, H. J. (2019) A survey on deep learning-based bearing fault diagnosis. Neurocomputing, 335, 327 – 335.*

[10] *Kang, Y., Chen, G., Wang, H., Pan, W., Wei, X. (2023) A new dual-input deep anomaly detection method for early faults warning of rolling bearings. Sensors, 23 (18), 8013.*

[11] *Shaalan, A., Mefteh, W., Frihida, A. M. (2024) Review on deep learning classifiers for faults diagnosis of rotating industrial machinery. Service Oriented Computing and Applications, 18 (4), 361 – 379.*

[12] *Wang, Z., Xu, Z., Cai, C., Wang, X., Xu, J., Shi, K., Zhong, X., Liao, Z., Li, Q. (2024) Rolling bearing fault diagnosis method using time-frequency information integration and multi-scale TransFusion network. Knowledge-Based Systems, 284, 111344.*

[13] *Wang, J., Zhao, Y., Wang, W., Wu, Z. (2024) Improving bearing fault diagnosis method based on the fusion of time–frequency diagram and a novel vision transformer. Journal of Supercomputing, 81 (1).*

[14] *Zhao, S., Xing, Y., Chen, X., Ou, Y., Zhang, W., Lin, B. (2020) A power metering pipeline fault warning method based on deep learning. In IOP Conference Series: Materials Science and Engineering, 740 (1), 012111. IOP Publishing.*

[15] *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.*

[16] *Zhu, W., Yu, C., Zhang, Q. (2022) Causal deep reinforcement learning using observational data. arXiv preprint arXiv: 2211.15355.*

[17] *Li, A., Luo, Y., He, Y., Cheng, Z., Wang, T. (2021) Fault diagnosis method of motor bearing based on deep transfer learning. In Proceedings of the IEEE China International Youth Conference on Electrical Engineering (CIYCEE), pp. 1 – 5. IEEE.*

[18] *Ren, H., Liu, S., Qiu, B., Guo, H., Zhao, D. (2024) A novel intelligent fault diagnosis method for bearings based on a multi-head self-attention convolutional neural network. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 38, e19.*

[19] *Chen, G., Qi, J., Gao, Y., Zhu, X., Dong, Z., Sun, Y. (2024) DGTRL: Deep graph transfer reinforcement learning method based on fusion of knowledge and data. Information Sciences, 658, 120019.*

[20] *Cao, H., Feng, F., Huo, J., Yang, S., Fang, M., Yang, T., Gao, Y. (2025) Model-based offline reinforcement learning with adversarial data augmentation. arXiv preprint arXiv: 2503.20285.*

[21] *Jiang, F., Liang, Q., Wu, Z., Kuang, Y., Zhang, S., Liang, J. (2024) A zero-cost unsupervised transfer method based on non-vibration signals fusion for ball screw fault diagnosis. Knowledge-Based Systems, 288, 111475.*