ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Edge-Cloud Collaborative Architecture for Real-Time Condition Monitoring and Intelligent Diagnosis of Offshore Platform Critical Equipment

Ying Zhang[1], *Tingting Wang*[1], *Xianlin Li*[1], *Xiaoyong He*[1] *and Yanlong Jiang*[1,*]

[1] Engineering Research and Design Department CNOOC Research Institute Ltd. No. 2 Building, No. 6 Courtyard, South Street of Taiyanggong, Chaoyang District, Beijing, 100028, China

SUMMARY: *The equipment on the offshore platform will be subjected to all kinds of harm from the sea, such as saltwater fog, high humidity, mechanical shock, etc. Equipment failure will result in a high-cost shutdown. The old way of periodic maintenance may be too frequent or too infrequent. A Three-Tier Edge-Cloud Collaborative Architecture for Online Condition Monitoring and Smart Fault Diagnosis of Offshore Platform Equipment is Proposed in this paper. The three layers of the system are: the terminal sensing layer for vibration signal collection; the edge computing layer, which uses a light-weight Lite-1D-CNN model for anomaly detection based on knowledge distillation; and the cloud analysis layer, which employs a ten-class fault classification model of multi-scale residual networks (MS-ResNet). An adaptive data transmission method based on the confidence of anomaly detection at the edge and task scheduling between edge and cloud nodes has also been introduced to reduce transmitted data by 58.3%. Based on the experiments in the bearing dataset at Case Western Reserve University, the edge model achieved an accuracy of 94.6% for binary classification and a latency of 6.8ms; the cloud model reached 97.1% accuracy for ten-class classification. The accuracy and weighted end-to-end latency of the collaborative edge-cloud model have been 96.2% and 44ms, respectively. The experiments on the bearing dataset at Paderborn University have also been verified for fault diagnosis of offshore platform equipment and achieved an accuracy of 86.5% in a different field.*

KEYWORDS: *edge-cloud collaboration; fault diagnosis; offshore platform; knowledge distillation; lightweight neural network; condition monitoring*

1 Introduction

Critical rotating equipment for offshore platforms in oil and gas production includes centrifugal pumps, gas compressors and generators. These machines operate in a harsh environment, such as a saline fog environment with high humidity and temperature fluctuations, and are subject to continuous vibration [1]. The above reasons have increased the rate of degradation, caused unexpected failures in equipment, led to production stoppages, resulted in safety hazards [2], and damaged the environment. Based on the above, time-based maintenance has led to either excessive or deficient maintenance, and thus the need for condition-based monitoring has arisen.

Deep learning is now being applied to the problem of mechanical fault diagnosis. Convolutional Neural Networks, Recurrent Neural Networks and Transformers can be used to

*jiangyl9@163.com
https://doi.org/10.65102/is2026949
Published: 30 April, 2026

learn discriminative features from raw vibration signals for bearing and gearbox fault diagnosis [3]. Parallel structures of CNN and LSTM networks can improve the accuracy of mechanical fault diagnosis under different conditions [4]. CNN-Transformer structures for multi-scale time-frequency representation have achieved state-of-the-art performance in mechanical fault diagnosis [5]. However, these computationally expensive models are difficult to run on low-power devices.

Edge computing performs local processing of signals and near-sensor inference to reduce the bandwidth and delay of communication [6]. Diagnostic models on edge devices can perform real-time fault detection without connecting to the cloud [7], but due to limited computing power, they are generally shallower than those used in the cloud.

To reduce the conflict between accuracy and real-time requirements, edge-cloud collaborative architectures have been proposed to offload the execution of coarse-grained tasks to edge devices and conduct complex analysis on cloud servers [8]. Scalable methods for heterogeneous fault diagnosis in distributed systems [9] and knowledge distillation for mapping complex models to efficient edge models for FPGA-based bearing diagnosis [10] have been studied. However, the current frameworks have not dealt with the offshore-specific limitations of limited communication capacity, multi-source environmental interference, and safety-critical real-time constraints.

The contributions of this study are as follows: First, a collaborative framework of the three-tier structure for real-time condition monitoring and fault diagnosis of offshore platform equipment is proposed, incorporating terminal sensing, edge computing and cloud analysis layers, and employing an adaptive data filtering mechanism based on edge computing detection confidence. Second, a lightweight one-dimensional convolutional neural network (Lite-1D-CNN) for edge-computing-based anomaly detection using knowledge distillation is proposed. Third, a multi-scale residual network (MS-ResNet) is used at the cloud layer for fine-grained ten-class fault classification. Fourth, large-scale experiments on the two bearing datasets verify the proposed architecture. Section 2 introduces the architecture and models; Section 3 shows experimental results; Section 4 lists implications and limitations; finally, Section 5 concludes the paper.

2 Materials and Methods

2.1 Data Description and Preprocessing

The first is the Case Western Reserve University (CWRU) Bearing Data Center dataset, and it has become a standard for bearing fault diagnosis research. A two-horsepower induction motor, a torque transducer, and a dynamometer are connected to it in the experiment. Test bearings are 6205-2RS JEM SKF bearings located at the drive end. Single-point faults are introduced at three locations by electro-discharge machining: the inner race, the outer race, and the rolling element; their fault diameters are 0.007", 0.014" and 0.021", respectively. Vibration signals are obtained from the four motor loads (0, 1, 2, and 3 HP) at a sampling rate of 12 kHz using drive-end accelerometers, and the corresponding shaft speeds are 1797, 1772, 1750, and 1730 RPM. A total of 10 health state categories are available.

The secondary dataset for cross-domain generalization testing is the Paderborn University Bearing Dataset; it includes vibration data of bearings with artificially induced and actual accelerated-life degradation faults under various rotational speeds and radial loads, sampled at a rate of 64 kHz. About 1200 samples in the inner-race and outer-race fault categories are extracted and preprocessed with the same windowing and STFT methods, but at different sampling frequencies.

To create an environment of strong background noise for offshore platforms in the simulation, add white Gaussian noise with a signal-to-noise ratio of 0-6 dB to the training data to improve its resilience to background vibration noise. The distribution of the datasets is as follows: CWRU training set 3360, validation set 720, test set 720; Paderborn cross-domain test set 1200.

2.2 Edge-Cloud Collaborative Architecture Design

The three layers in the new structure are: the top-level terminal sensing layer, the mid-level edge computing layer, and the bottom-level cloud analysis layer, as shown in Figure 1. Accelerometers and temperature sensors in the terminal sensing layer will be employed to obtain raw vibration signals from critical rotating equipment, segment these signals into fixed-size time windows of 1024 points, and then transmit them to the edge computing node.

The first stage of intelligent processing is the edge computing layer. Each edge computing node in this layer will use the Lite-1D-CNN model for real-time binary anomaly detection and return the corresponding confidence score. An adaptive data transmission mode will be used to reduce the volume of data sent to the cloud; only those data packets that are considered normal with a high confidence level (>0.95) will be sent, and these will be represented by a brief four-feature statistical summary. If the confidence is too low or an anomaly occurs, then the original data will be transmitted. In this way, a smaller amount of communication bandwidth will be required in the offshore area because a satellite or a low-bandwidth maritime network will be used.

The MS-ResNet deep classification model in the cloud analysis layer is used for fine-grained ten-class fault diagnosis, long-term trend analysis and periodic model management. The cloud model will classify the fault type, position and degree of abnormality of the sample. A model update protocol retrains the cloud model and compresses the updated knowledge into the Lite-1D-CNN student model, distributes the updated parameters to edge nodes at 24-hour intervals or after accumulating 500 newly identified anomalous samples, whichever comes first.

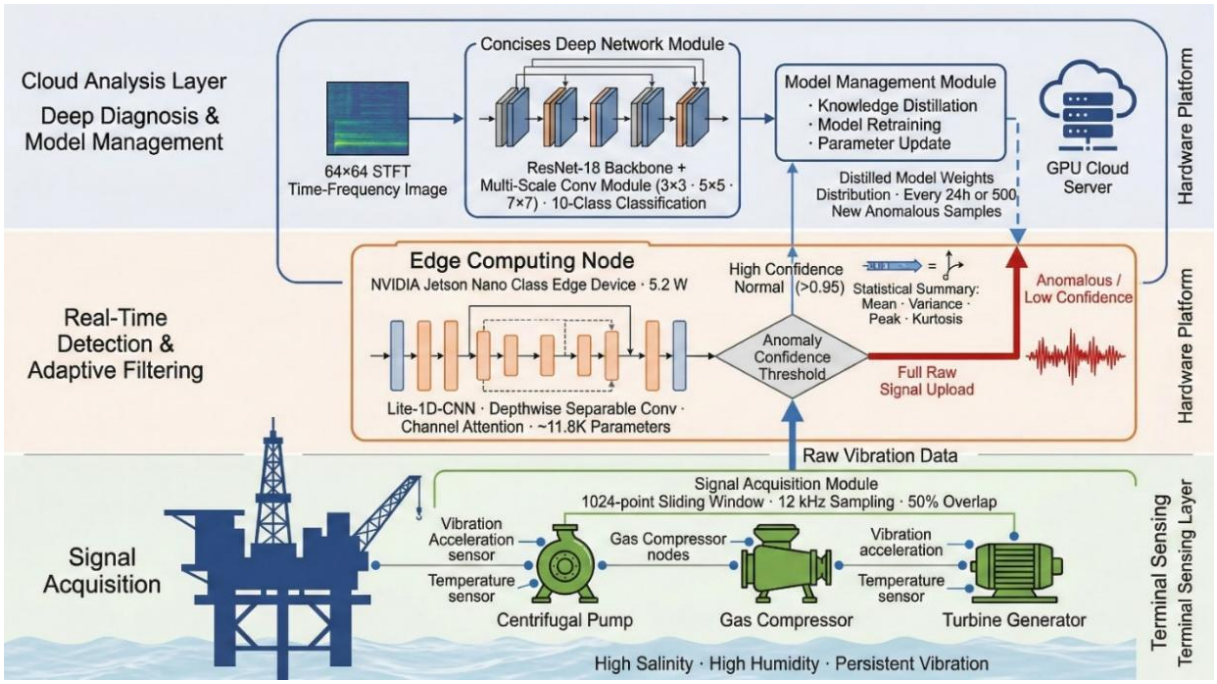

Figure 1: *Overall Framework of the Three-Layer Edge-Cloud Collaborative Architecture.*

2.3 Edge-Side Lightweight Anomaly Detection Model

Lite-1D-CNN is a relatively light 1D CNN that is feasible for real-time anomaly detection. The four one-dimensional convolutional layers in the architecture are followed by a global average pooling layer and a fully connected output layer. Depthwise separable convolutions replace the standard convolutions in the second and third layers to reduce computational load while maintaining the representative ability. A simple channel attention mechanism is used after the third convolutional layer; global average pooling is applied across the temporal dimension to obtain a channel descriptor vector, which is then processed by a two-layer bottleneck with a reduction factor of 4 and sigmoid gates. The output layer is a two-class probability distribution (normal/anomalous), and the confidence score is used to decide whether to adaptively transmit.

Lite-1D-CNN is trained using knowledge distillation with the cloud-side MS-ResNet as the teacher network. Soft probability outputs from the teacher are combined with hard binary labels under a joint distillation loss: L = $\alpha \cdot$ L_soft + (1 - α) $\cdot$ L_hard, with $\alpha = 0.7$. Adam optimizer is used for training; learning rate is 5e-4 and batch size is 128 over 80 epochs. INT8 post-training quantization is used for further compression.

As shown in Table 1, the final model has 11.8K parameters, 1.9M FLOPs, and 14.2 KB model size after quantization. Compared with a one-dimensional MobileNetV2 variant (268.4K parameters, 42.7M FLOPs, 312.5 KB), the proposed model has reduced both parameters and FLOPs by more than 95% and is thus suitable for deployment on power-constrained offshore edge devices.

Table 1: *Structural Comparison of Edge-Side and Cloud-Side Models.*

Model	Task	Input	Layers	Params	FLOPs	Size (INT8)
Lite-1D-CNN (Edge)	Binary	1×1024	4 Conv+GAP+FC	11.8K	1.9M	14.2 KB
MobileNetV2 (Edge)	Binary	1×1024	17 Bottleneck+FC	268.4K	42.7M	312.5 KB
Standard 1D-CNN (Edge)	Binary	1×1024	4 Conv+FC	35.6K	5.8M	41.3 KB
MS-ResNet (Cloud)	10-class	64×64	ResNet-18+MS	11.24M	1.83G	—
ResNet-18 (Cloud)	10-class	64×64	18 layers	11.17M	1.82G	—

2.4 Cloud-Side Deep Fault Classification Model

MS-ResNet in the cloud uses a modified ResNet-18 and has added a multi-scale convolutional feature fusion module. The input is a 64x64 time-frequency image obtained through short-time Fourier transform. After the initial convolutional stem and the first two residual blocks, a multi-scale extraction module is used for the feature map, and at the same time, parallel convolutional kernels of sizes 3×3, 5×5 and 7×7 are employed to obtain features at different frequency resolutions. Parallel outputs are concatenated and fused by a 1×1 convolution. Next, the features are passed through the rest of the residual blocks, global average pooling, and the final fully connected layer, and then the probabilities of the features belonging to the ten health state categories are outputted.

The classification targets are one normal state and nine fault states, which fall into three categories of fault locations: inner race, outer race, and rolling element; and these faults have three severity levels: 0.007, 0.014, and 0.021 inches. The Adam optimizer is used for training with an initial learning rate of 1e-3, a cosine annealing learning schedule over 100 epochs, a batch size of 64, and an early stopping patience of 15 epochs. The combined loss function is a sum of cross-entropy loss and center loss, and the weight of the latter is 0.003. It will improve the discrimination ability of the model for fault states with similar characteristics, particularly those of rolling element faults at different severities and the normal state. Other indices are overall accuracy, precision, recall, F1 score, confusion matrix and inference time.

3 Results

3.1 Edge-Side Real-Time Anomaly Detection Performance

A distillation-trained Lite-1D-CNN model achieved a binary classification accuracy of 94.6%, an F1-score of 0.941, and a recall rate of 95.2%; it could identify most of the faulty samples correctly with a low false alarm rate that was still suitable for adaptive transmission.

Noise robustness is tested in the presence of additive white Gaussian noise at five SNR levels (-2, 0, 2, 4, and 6 dB), as shown in Figure 2. At an SNR of 6 dB, the accuracy and F1-score of the model are 95.1% and 0.946. Accuracy gradually decreases to 93.2% at SNR = 0 dB and 90.3% at SNR = −2 dB. Without distillation, the model's accuracy is consistently 2.5-3.5 percentage points lower at all noise levels; thus, it can be confirmed that knowledge distillation improves generalization under distributional shifts. Despite having over 22 times more parameters, MobileNetV2 only marginally outperforms Lite-1D-CNN at the highest SNR (94.7% vs. 95.1%) and drops to 89.1% at SNR = -2 dB. The baseline 1D-CNN without distillation or attention mechanisms performs the worst at all noise levels and is 86.7% at SNR = -2 dB.

The latency of single-sample inference in the simulated Jetson Nano environment is around 6.8ms, and thus it meets the real-time requirement for online monitoring at a 12kHz sampling rate.

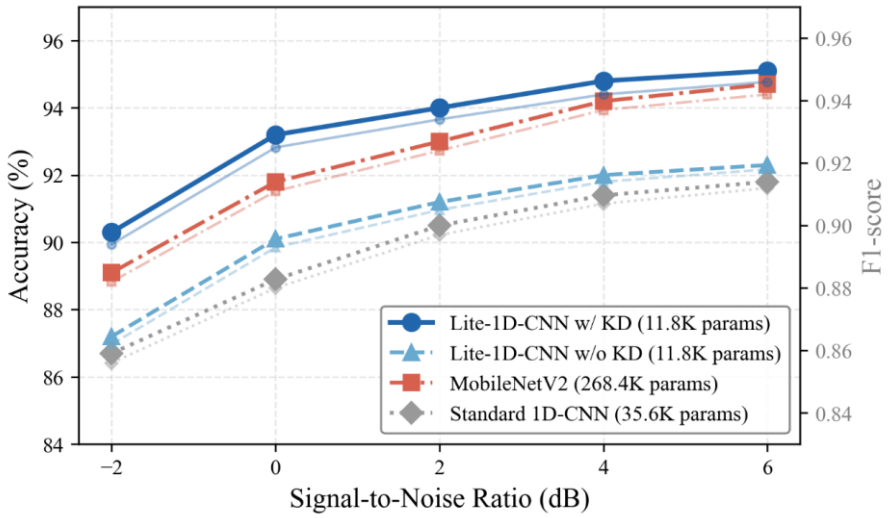

Figure 2: Performance of Edge-side Anomaly Detection under Different Noise Levels.

3.2 Cloud-Side Fault Classification Accuracy

MS-ResNet achieved a top-ten classification accuracy of 97.1% on the CWRU test set and a macro-averaged F1-score of 0.965. The normalized confusion matrix in Fig. 3 shows that most fault types are recognized above 96.5%. Inner race faults at all three severities have achieved accuracy rates of 97.5%-98.8% due to clear spectral signatures, and outer race faults have achieved 96.9%-98.5%.

Misclassification occurs most frequently in the category of rolling element faults. Ball-0.007 has the lowest recognition rate (93.8%), with 4.2% incorrectly identified as normal and 1.3% as Ball-0.014; therefore, it is difficult to distinguish between early-stage rolling element faults and normal vibrations. Outer race faults at 0.007 inches exhibit minor cross-severity confusion (2.1% misclassified as OR-0.014). All other off-diagonal elements of the confusion

matrix are less than 1.5%, and the multi-scale feature fusion module has achieved good fault-type discrimination.

The classification accuracy at different load levels is as follows: 97.8% for 0 HP, and it gradually decreases to 97.3%, 96.9%, and 96.1% at 1, 2, and 3 HP, respectively, because the increased vibration energy and frequency spread make the fault characteristics less distinct. MS-ResNet is approximately 1.4% more accurate than the previous single-scale ResNet-18 model overall.

Figure 3: Normalized Confusion Matrix of MS-ResNet for Ten-Class Fault Classification.

3.3 Edge-Cloud Collaborative System Evaluation

The performance of the proposed integrated edge-cloud collaborative framework is compared with two different baselines: edge-only deployment (Lite 1D CNN for direct classification without fine-grained capability) and cloud-only deployment (all data transmitted to MS ResNet without edge filtering). The indicators of the evaluation are as follows: Table 2.

The proposed collaborative framework has reduced cloud-bound data transmission by 58.3% compared with the cloud-only deployment. At a confidence threshold of 0.95, 78.2% of all samples are classified as confident normal at the edge and are presented as four-dimensional statistical summaries; 21.8% of all samples are transmitted in full, generating 20.3 MB per 1000 samples, and 48.6 MB are generated for every 1000 samples in the cloud-only deployment.

A new cooperative framework will reduce the total delay through edge-side pre-processing. Edge-resolved samples have an average latency of 9.2ms, and cloud-routed samples are as high as 168ms due to network round-trip times. The weighted system-wide average latency is about 44 ms, which is an improvement of 80.4% over the cloud-only deployment (225 ms), and it is below the 50 ms engineering threshold for real-time monitoring of rotating machinery.

The proposed collaborative framework has a total diagnostic accuracy of 96.2% and is 0.9

percentage points lower than that of the cloud-only deployment (97.1%). Approximately 2.8% of genuinely abnormal samples are classified as high-confidence normal at the edge and thus avoided cloud-side diagnosis. Although its system miss rate of 3.8% is higher than that of the cloud-only deployment (2.9%), it is still relatively low compared with the edge-only deployment (12.7%). The accuracy loss is deemed acceptable for the offshore platform, and other-shore latency and bandwidth reduction are required.

Table 2: Performance Comparison of Three Deployment Strategies.

Metric	Edge-Only	Cloud-Only	Proposed
Fault classification accuracy (10-class)	87.3%	97.1%	96.2%
Average end-to-end latency	9.2 ms	225 ms	44 ms
Data transmission (per 1000 samples)	0 MB	48.6 MB	20.3 MB
Bandwidth reduction vs. Cloud-Only	—	Baseline	58.3%
Missed fault rate (false negative)	12.7%	2.9%	3.8%
Edge device power consumption	5.2 W	—	5.2 W
Cloud GPU utilization	—	100%	37.4%

3.4 Comparison with Benchmark Methods and Cross-Domain Validation

The multi-dimensional comparison of the deployment strategies is shown in Figure 4. The proposed edge-cloud collaborative system has achieved a 96.2% accuracy rate at a latency of 44ms, thus meeting the 50ms real-time requirement. The cloud-based ResNet-18 and MS-ResNet models have achieved an accuracy of 97.1% at 225 ms and 232 ms latency, and thus meet the real-time requirements. Edge-based MobileNetV2 and 1D CNN models have lower latency but reduced accuracy: 91.5% and 87.3% at 12.6 ms and 9.2 ms latency, respectively. The proposed model is a trade-off between accuracy and latency that is relatively small, and it lies on the Pareto front.

For cross-domain generalization, the proposed MS-ResNet model trained on the CWRU dataset is tested on the Paderborn University bearing dataset without fine-tuning and achieves an accuracy of 86.5%, a decrease of 10.6% compared to the CWRU-trained model. The baseline ResNet-18 model has achieved an accuracy of 83.2% and thus, the multi-scale module provides a gain of 3.3%. The cross-domain performance shows that the learned features are somewhat specific to the measurement environment at CWRU, such as sensor locations, bearing structures, noise characteristics, etc.

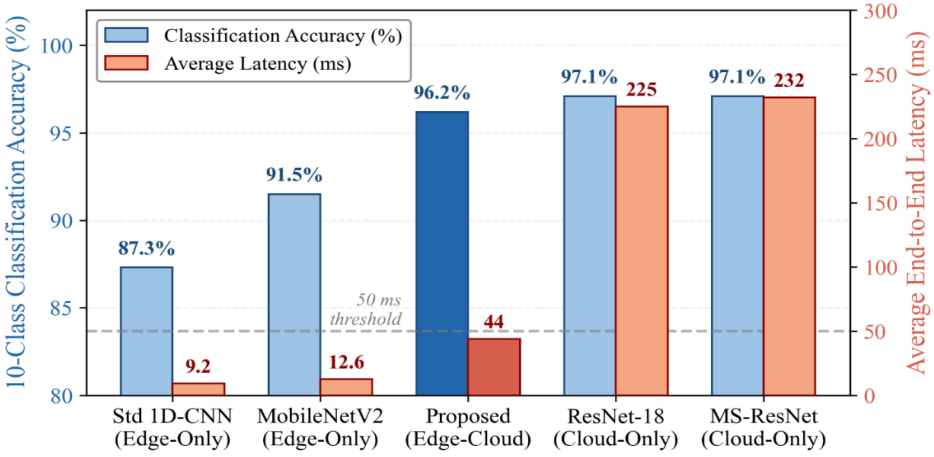

Figure 4: Comparison of Classification Accuracy and Latency under Deployment Strategies.

4 Discussion

4.1 Confidence Threshold and Latency-Accuracy Trade-off

The confidence threshold of the edge layer is used to decide whether to perform local processing or forward to the cloud, and thus trades off latency for accuracy. Table 3 shows the summary of four different levels of confidence. When set to 0.85, 88.4% of the samples are locally processed with a very low latency of 34ms, but at the cost of a high missed fault rate of 5.5% and the lowest accuracy of 94.5%. At 0.99, 43.7% of the samples are forwarded to the cloud to maximise accuracy (96.8%) but increase latency to 103 ms. A threshold of 0.95 is selected; thus, the accuracy is 96.2%, latency is 44ms, and the rate of missed faults is 3.8%. In safety-critical applications where an undetected fault would be fatal, a higher threshold can be set at the cost of reduced bandwidth and latency [11].

Knowledge Distillation Hyperparameters Affect Edge Model Quality. At the temperatures of 2 and 8, the teacher output smoothing and discriminative sharpness are unbalanced, so T=4 is selected. A soft-hard label weighting of $\alpha=0.7$ is better than both the pure soft label option ($\alpha=1.0$, 93.1% accuracy) and the predominantly hard label option ($\alpha=0.3$, 93.8%); thus, it can provide inter-class structural information that supplements the categorical signal of hard labels, and is consistent with recent adaptive distillation work in resource-constrained diagnostic scenarios [12].

Table 3: Sensitivity Analysis of Edge-side Confidence Threshold Settings.

Threshold	Edge (%)	Cloud (%)	Accuracy (%)	Latency (ms)	Missed Fault (%)
0.85	88.4	11.6	94.5	34	5.5
0.90	83.7	16.3	95.4	38	4.6
0.95*	78.2	21.8	96.2	44	3.8
0.99	56.3	43.7	96.8	103	3.2

4.2 Applicability and Practical Considerations for Offshore Platforms

The experiment's data are in a lab setting; offshore platforms, however, face various non-stationary noise that is not included in the lab data, such as tidal fluctuations, wave-induced structural vibrations, machine coupling noises, and fluid pulsation. They are not the same as the additive white Gaussian noise for data augmentation. Although noise injection at 0 dB SNR partially simulates environmental noise, it does not fully represent the actual conditions of marine noise.

Another deployment constraint that communication infrastructure poses is that the platforms will have access to satellite communication with a bandwidth of less than 512 kbps for remote platforms, and nearshore platforms will be connected to 4G or 5G maritime base stations. Although the new filter method reduces the amount of transmission by 58.3 per cent, it has not been tested in a disconnection-prone or data-loss environment. The power consumption of the proposed edge model at 5.2W is still within the allowed range for industrial edge computing devices and auxiliary power supplies of offshore platforms [13].

4.3 Limitations and Future Directions

However, the following deficiencies must also be pointed out. First, the datasets that will be used have relatively simple operating conditions, fault geometries and noise environments compared with those of offshore platforms. Secondly, the current model is only for bearing faults, and monitoring offshore platforms will need to include gear box failures, pump seal

failures, shaft misalignment, etc. Thirdly, the model update process is based on fixed intervals and thus cannot promptly respond to changes in the operating environment. Finally, with a cross-domain model accuracy of 86.5%, it can be seen that there is still a considerable gap in performance, and the model may have overfitted to the acquisition-specific features.

There are some promising fields for future study. For example, the model can be used to test and validate the offshore platforms by integrating it with the SCADA system of the platforms [14]. Frameworks for federated learning can be employed to jointly optimise the model without accessing sensitive operational data[15]. A digital twin of the system can be built to conduct early warning predictions of failures based on data before actual damage occurs [16]. The scope of the model can be expanded to cover all faults and equipment in order to improve the generalizability of the architecture for managing offshore platforms.

5 Conclusion

This paper has presented a three-layer edge cloud collaborative architecture for the real-time condition monitoring and smart fault diagnosis of critical equipment at offshore platforms. Perform experiments on the CWRU bearing dataset to validate the performance of the introduced edge cloud collaborative architecture. Based on the above results, the edge-side Lite-1D-CNN model has an accuracy of 94.6% for binary anomaly detection and an inference speed of 6.8 ms per sample on the CWRU bearing dataset after INT8 quantization, with 11.8K parameters and a model size of 14.2 KB. Knowledge distillation of the cloud-side model has improved the performance of the edge-side Lite-1D-CNN model by 2.9% compared to the Lite-1D-CNN model without knowledge distillation. The MS-ResNet model in the cloud achieves 97.1% accuracy on the ten fault categories of the CWRU bearing dataset. The performance of the MS-ResNet model is about 1.4% higher than that of ResNet-18 on the same data. When building an end-to-end latency-reducing collaborative architecture for edge clouds, it can be lowered to 44ms from 225ms in the cloud-side model; at the same time, about 58.3% of the transmitted data has been reduced, and the overall diagnostic accuracy has dropped slightly from 97.1% in the cloud-side model to 96.2% in the edge cloud collaborative architecture.

Based on the above studies, some shortcomings have been identified. For example, according to cross-domain evaluation results in the Paderborn University bearing dataset, the achieved accuracy of 86.5% is 10.6 percentage points lower than that possible. This is a sign that there may be some difficulties in generalizing a model built on a single set of laboratory conditions to other situations, particularly off-shore platforms. However, based on the above study, it is feasible to maintain an accuracy of over 96 per cent, reduce the response delay to less than 50 milliseconds, and significantly decrease communication bandwidth. Therefore, it can be seen that the proposed architecture serves as the foundation for constructing an intelligent condition monitoring system in an offshore platform.

References

[1] Maurya, M., et al., Intelligent fault diagnostic system for rotating machinery based on IoT with cloud computing and artificial intelligence techniques: a review: M. Maurya et al. Soft Computing, 2024. 28(1): p. 477-494.