INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING
ISSN: 2229-8649 e-ISSN: 2180-1606
VOLUME 23, ISSUE 2, 2026, 13568 - 13582
DOI: https://doi.org/10.15282/ijame.23.2.2026.8.1027

RESEARCH ARTICLE

Cross-machine reliability and fault diagnosis using correlation alignment, feature alignment, and enhanced deep extreme learning machine

Muhammad Amir Zikry Harun*, Muhammad Firdaus Isham, Mohd Syahril Ramadhan Saufi, Wan Aliff Abdul Saad, and Muhammad Danial Abu Hasan

Fakulti Kejuruteraan Mekanikal, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

Abstract - Reliability and efficiency are vital components in modern manufacturing systems, especially for rotating machinery that operates continuously under a variety of environments. Recently, there has been growing interest among researchers in developing and deploying a cross-machine fault-diagnosis system in real-world industrial settings. It allows engineers to perform fault diagnosis across multiple machines without having to build a new intelligent model each time. Unlike most studies that utilized complex transfer learning architectures, this paper proposes a simple and efficient cross-machine bearing fault diagnosis framework based on correlation alignment and deep extreme learning machine (DELM). It aligned time-domain statistical features from two datasets, Case Western Reserve University and the experimental dataset, to reduce the domain gap between them and improve generalization performance. The aligned features were classified using a DELM model. The experiment yielded reliable cross-machine generalization, with an average accuracy of 91.27%. These findings underscore the model's capability to offer a simple, effective, and low-computational-demand solution for cross-machine applications, making it suitable and practical for real industrial use.

Article History
Received : 14 August 2025
Revised : 26 February 2026
Accepted : 29 April 2026
Published : 30 June 2026

Keywords
Vibration analysis
Bearing fault diagnosis
Cross-machine
Deep learning
Extreme learning machine
Correlation alignment

1. Introduction

Operational efficiency and reliability are critical for industries to meet customer demands. Manufacturing plants often face challenges in maintaining operational efficiency amid rapidly increasing customer demand. Consequently, engineers have developed multiple approaches to address the issue, with preventive maintenance (PM) becoming the dominant and most efficient strategy in many industries. It has been reported that machine failures account for 15-60% of manufacturing costs and up to 50% of total production costs in heavy industries [1]. PM aims to reduce machine downtime, allowing manufacturers to keep their machinery in good condition and meet operational demands [2]. Also, this approach is utilised to ensure high-quality outputs and prolong machinery lifespan in a highly competitive market [3]. PM involves scheduled maintenance, including inspection, lubrication, and component replacement to ensure machines can run continuously [4]. In real-world industrial scenarios, predictive maintenance, a subset of PM, has gained attention in recent years for employing a data-driven maintenance strategy that bases maintenance actions on machine health rather than fixed time intervals. Among the data types introduced to facilitate predictive maintenance, including temperature, pressure, and acoustic emission, vibration-based monitoring has been gaining popularity in rotating machinery applications due to its high sensitivity for early fault detection, making it a suitable approach for real-world industrial applications [5]. Although all approaches mentioned above are highly efficient, developing a fault detection and diagnosis system that processes large amounts of vibration data often requires a machine learning (ML) model. As technology continues to advance and critical machinery systems become more complex, ML has become an essential tool, widely used across many industries to improve maintenance strategies. In particular, deep learning (DL), a subset of ML and artificial intelligence (AI), has emerged as a powerful tool, gaining attention for its ability to handle large datasets while maintaining high accuracy in fault diagnosis. DL draws on the human brain, mimicking the activity of its neurons as the foundation of its operation [6]. DL uses multiple layers of artificial neural networks to extract relevant information, allowing it to handle complex tasks such as image recognition and natural language processing [7]. In intelligent maintenance systems, DL can perform well in classifying fault conditions in rotating machinery due to its strong diagnostic capability. However, the DL-based approach often relies heavily on large, high-quality labelled datasets for it to perform at its best [8]. DL models also typically have a large number of parameters, which requires substantial computational resources [9].

To address this issue, Huang G. B., a researcher, developed the extreme learning machine (ELM) model. ELM is introduced as a faster and simpler alternative to data-driven learning models. Unlike DL, ELM is based on a single-layer feedforward network that randomly assigns input weights and biases yet still produces reliable outputs [10]. ELM offers faster data processing without backpropagation, making ELM an ideal option for real-time applications [11], [12]. In general, industries require high-speed AI models with low computational requirements. Additionally, ELM performance can be further improved by incorporating a kernel function. A kernel allows the model to handle complex, nonlinear, high-dimensional data more effectively. Kernel functions implicitly map input features into higher-dimensional spaces to capture patterns that would otherwise be difficult for a standard linear model to capture [13]. This improvement makes ELM an ideal candidate for tasks like fault diagnosis in industrial settings, where both speed and accuracy are essential. DL and ELM both have strengths and weaknesses when deployed in industrial fault diagnosis systems. As a result, researchers have proposed integrating both models into a unified framework. The integration leverages DL's ability to

*CORRESPONDING AUTHOR | M. A. Z. Harun | ✉ muhammadamirzikry@graduate.utm.my

handle complex data and ELM's ability to provide fast outputs, as in fault diagnosis [14]. The integration of both models is named deep extreme learning machine (DELM). In recent years, there has been an increased interest in developing AI models for cross-machine fault diagnosis applications [15]. Cross-machine is a practical solution that aims to reduce the need to collect and label large datasets across multiple machines, a process that is often difficult and costly [16]. The performance of DELM has been proven to be reliable in single-machine fault diagnosis [14]. While the integration of deep learning and extreme learning machines has been explored in fault diagnosis systems, all research papers mainly focus on single-machine applications. Existing studies on cross-machine applications typically rely on heavy, complex transfer learning or domain adaptation frameworks, resulting in high computational costs. In contrast, the deep causal factorisation network (DCFN) introduces causal representation learning, in which vibration data from two machines are decomposed into causal fault-related features, removing domain-related factors that can affect generalisation capability [17]. However, DCFN relies heavily on deep architecture and multiple learning modules, increasing model complexity. Apart from that, the deep domain adaptive transfer network (DDATN) [18] focuses on aligning both marginal and conditional feature distributions between source and target domains using deep parameter-shared networks and pseudo-label learning. However, reliance on iterative backpropagation and dynamic domain adaptation may prolong training time.

A physics-informed unsupervised domain adaptation (PIUDA) framework [19] was proposed, incorporating both fault-related physical knowledge and spectral-based indicators into the domain adaptation process. On the other hand, a multi-task self-supervised transfer learning framework (MTSTLF) [20] learns transferable fault features across multiple self-supervised masking tasks and across machines via multi-perspective feature alignment. Although these frameworks have successfully demonstrated strong cross-machine capability for bearing fault applications, they rely heavily on complex model architectures, which in turn impose high computational demand. Their complexities and training costs limit their practicability in industrial environments. Consequently, this study proposes a cross-machine bearing fault diagnosis framework using correlation alignment and DELM. The main contribution of this study is the development of a simple, computationally efficient method for classifying bearing fault conditions across different machines without relying on deep, complex architectures. This study focused on diagnosing four bearing conditions: healthy, ball fault, inner-race fault, and outer-race fault. The diagnosis was done by analysing vibration data collected from two sources. The first source was the Case Western Reserve University (CWRU) online bearing dataset [21], while the second was an independently collected experimental dataset. Time-domain statistical features, such as mean, peak-to-peak, and others, were extracted from the raw vibration data to serve as the basis for model training and testing. To address the data distribution mismatch across machines, correlation alignment (CORAL) was applied to align the second-order statistics. The aligned features were then classified using the DELM model, comprising two ELM-based Autoencoder layers followed by an ELM classifier. The model was trained solely on source-domain features, and its cross-machine generalisation capability was primarily tested using generalised target-domain features. The proposed framework's performance was assessed using multiple classification metrics, including accuracy, precision, recall, and F1 score, to provide a comprehensive evaluation of its diagnostic effectiveness.

1.1 Deep Learning

DL is a powerful and versatile tool in AI and ML. The architecture of a common DL is illustrated in Figure 1. It employs multi-layered artificial neural networks that can automatically learn patterns from the input data without human intervention. This makes DL different from traditional machine learning models, which require manual feature extraction before training [22]- [23]. The ability to extract relevant information from raw data enhances the model's adaptability and accuracy. As a result, most researchers prefer to employ a DL algorithm to perform complex tasks that may not be well-suited to traditional machine learning models, such as image recognition, natural language processing, and fault diagnosis [24]. Apart from that, DL is also commonly known for its ability to handle large datasets. As technology continues to advance, the volume of data to be analysed keeps growing. Thus, DL is the preferred model for handling large datasets while maintaining the accuracy and reliability of its diagnostic results [25]. Furthermore, a variety of DL architectures can be used for different purposes, such as a Recurrent Neural Network for time-series and sequence data, and an Autoencoder for feature extraction and data reconstruction. However, among all DL architectures, Convolutional Neural Networks have been widely used in industrial applications, especially for analysing and recognising time-frequency images for fault diagnosis [26]. The architecture of a common DL model includes an input layer, a hidden layer, and an output layer. However, in practice, it can be designed with multiple hidden layers to perform well on complex tasks. Each hidden layer progressively extracts higher-level features from the input data. The final hidden layer, also known as a fully connected layer, is used for classification. Additionally, in DL architectures, activation functions such as ReLU (Rectified Linear Unit) are used to handle nonlinear data. This allows the model to capture and learn the complex patterns within the data [27].

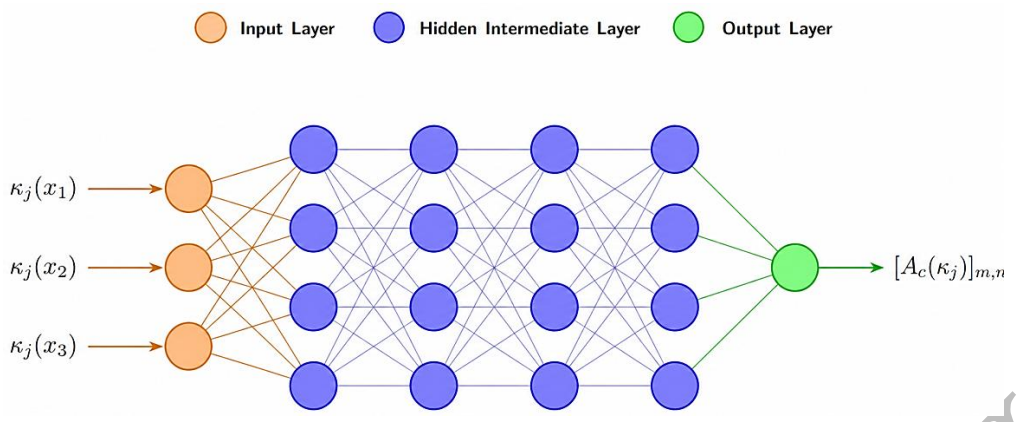

Figure 1. Deep learning architecture

1.2 Extreme Learning Machine

ELM is a fast, simple, and efficient learning model developed for single-hidden-layer feedforward neural networks. ELM is unique among AI models because it eliminates backpropagation during training by randomly initialising input weights and biases and computing the output weights analytically. This ability helps reduce training time [28]. The simple approach used in ELM reduces computational demand while achieving high accuracy in diagnostic tasks. This makes ELM an ideal option for industrial real-time applications and scenarios involving limited datasets. The architecture of ELM, as illustrated in Figure 2, consists of three main layers: an input layer, a hidden layer with randomly assigned weights and biases, and an output layer whose weights are computed directly. Moreover, ELM's performance can still be improved by integrating the kernel function. A kernel is used to handle complex, nonlinear tasks. A kernel can map the input data into a higher-dimensional space using its matrices, facilitating better separation and classification without relying on random hidden-layer mappings [29], [30]. The integration of kernel functions and ELM, also known as KELM, is used in many industrial applications, such as power transformer diagnostics and cross-domain fault detection in aerospace, due to its strong generalisation and stability [31].

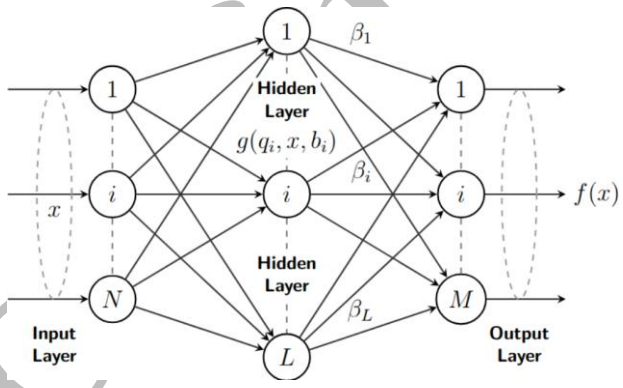

Figure 2. ELM architecture

1.3 Correlation Alignment

CORAL is a common domain adaptation (DA) technique that aims to minimise the data distribution gap between the source and the target domain. In practical fault diagnosis applications, vibration data collected from different machines often exhibit distinct characteristics due to varying operating conditions, sensor placement, and structural design. To address the issue, CORAL offers a strategic solution by aligning the second-order statistics, which are the covariance matrices. By reducing statistical differences across two or more domains, CORAL enables the AI model to learn and extract more fault-related features than domain-focused ones, thereby improving overall generalisation [32]. One of the main advantages of using CORAL is that it performs well in both supervised and unsupervised learning applications. In real-world scenarios, labelled data from the target domain is typically limited due to the time-consuming and high operational costs. CORAL can be effectively applied in such cases because it does not require access to target-domain labels for alignment. This characteristic makes CORAL the preferred solution for cross-machine tasks, where collecting and labelling data from the target domain can be impractical. In addition, CORAL is also a method that does not require extensive hyperparameter tuning or iterative optimisation [33], offering simple integration into existing machine learning frameworks [34]. Due to its simplicity and effectiveness in bridging gaps across multiple data distributions, CORAL has been widely adopted in cross-machine applications.

Compared with other DA techniques, such as maximum mean discrepancy (MMD), adversarial DA, and contrastive learning-based methods, CORAL has been the preferred solution due to its simplicity and computational efficiency. MMD-based methods, for instance, rely heavily on kernel functions to measure distributional differences. Incorrect kernel selection and parameter configuration may negatively affect performance, especially in deep-architecture models.

Adversarial DA methods, such as domain-adversarial neural networks (DANN), have been widely adopted into industrial settings due to their reliable performance in learning fault-related features and avoiding domain-related factors through adversarial training [35]. DANN, however, introduced additional discriminator networks and gradient reversal mechanisms that may lead to training instability. It also requires careful hyperparameter tuning to perform at its best. More recently, contrastive learning has gained attention for its promising ability to learn transferable representations. It is done by enforcing instance-level or feature-level similarity across domains. Despite its effectiveness, this method requires careful design of positive-negative pair construction and the use of an additional optimiser, thereby increasing model complexity. In contrast, CORAL aligns second-order statistics via covariance matching and applies a linear transformation directly to the feature space, without adding any additional trainable components [36]. In practice, CORAL can be implemented in multiple ways, including common steps like whitening and recolouring. The first step of CORAL is to compute the covariance matrices of the source domain (C_s) and the target domains (C_t). The covariances are regularised and used to derive a transformation matrix (A) to map the target-domain features into the source-domain feature space. By following these steps, the statistical properties of the target features can align with those of the source domain, effectively reducing domain discrepancy. The mathematical formulation of CORAL is presented in Equations (1)-(4).

$$C_s = cov(X_s) + I \tag{1}$$

$$C_t = cov(X_t) + I \tag{2}$$

$$A = \sqrt{C_s} \cdot \left(\sqrt{C_t}\right)^{-1} \tag{3}$$

$$X_t^{Aligned} = (X_t - \mu_t) \cdot A + \mu_t \tag{4}$$

where X_s and X_t represent feature matrices of the source and target domains, C_s and C_t are the covariance matrices, I is the identity matrix, μ_t is the mean of the target domain distribution, and A represents the transformation matrix.

2. Materials and Methods

A structured workflow, as illustrated in Figure 3, was developed in this study to ensure the proposed CORAL-DELM could be implemented smoothly to achieve the research objectives. The workflow consisted of four main phases. The first phase focused on data collection and preparation using two vibration datasets, one from the CWRU online bearing dataset (source domain) and the other from an experimental dataset (target domain). Phase 1 began by acquiring raw vibration data from both sources. The CWRU dataset consisted of vibration signals collected under controlled laboratory conditions with artificially induced bearing faults. It consisted of four classes: healthy, ball fault, inner-race fault, and outer-race fault. All faults were induced with a constant defect diameter of 0.007 inches. Furthermore, the motor was run at a constant speed of 1797 RPM under no-load conditions. The sampling frequency utilised in data collection was 12 kHz. Figure 4(a) displays the experimental setup of the CWRU bearing dataset, and further details regarding the test rig configuration are available from the official CWRU dataset documentation [21]. On the other hand, the experimental vibration signals were obtained using a laboratory test rig, the Machinery Fault and Rotor Dynamics Simulator (MFS-RDS), designed by SpectraQuest. To match the classes in the CWRU dataset, the experiment was also conducted to collect data for four similar classes. The setup configuration is demonstrated in Figure 4(b). The motor was run at a constant speed of 1800 RPM with no load. The sampling frequency utilised in the experiment was 25.6 kHz.

The raw vibration signals were then segmented into fixed-length samples at the sampling rate and motor speed for each dataset. Each sample was aimed to represent one full shaft rotation, ensuring uniform patterns across both domains. For the CWRU dataset, a motor speed of 1797 rpm and a sampling rate of 12 kHz were used, yielding 400 data points per sample. On the other hand, the experimental dataset was operated at a motor speed of 1800 rpm with a sampling rate of 25.6kHz, resulting in 853 data points per sample. Segmenting the raw data was an essential step in a fault diagnosis framework, as it highlighted the relevant information for analysis and accelerated processing. By ensuring that each sample captures a single mechanical cycle, this segmentation step improved the consistency of feature extraction between the two domains, thereby enhancing cross-machine diagnostic capability. Moreover, the detailed configuration of data points for both datasets is summarised in Table 1. For each dataset, 400 signal samples were obtained and processed through the feature-extraction and normalisation steps. Feature extraction facilitated the capture of key information in each sample. In this study, the focus was on extracting time-domain statistical features, as listed in Table 2. Initially, a set of statistical features was extracted from the raw signals. These features were then examined empirically and categorised according to their ability to distinguish among different bearing fault conditions. Features that showed observable differences between fault conditions in both datasets were retained, while redundant features lacking distinguishing characteristics were excluded. Therefore, 10 time-domain statistical features were chosen because they provide sufficient diagnostic information while avoiding unnecessary computational overhead. These 10 time-domain features were then normalised before proceeding to the feature generalisation step.

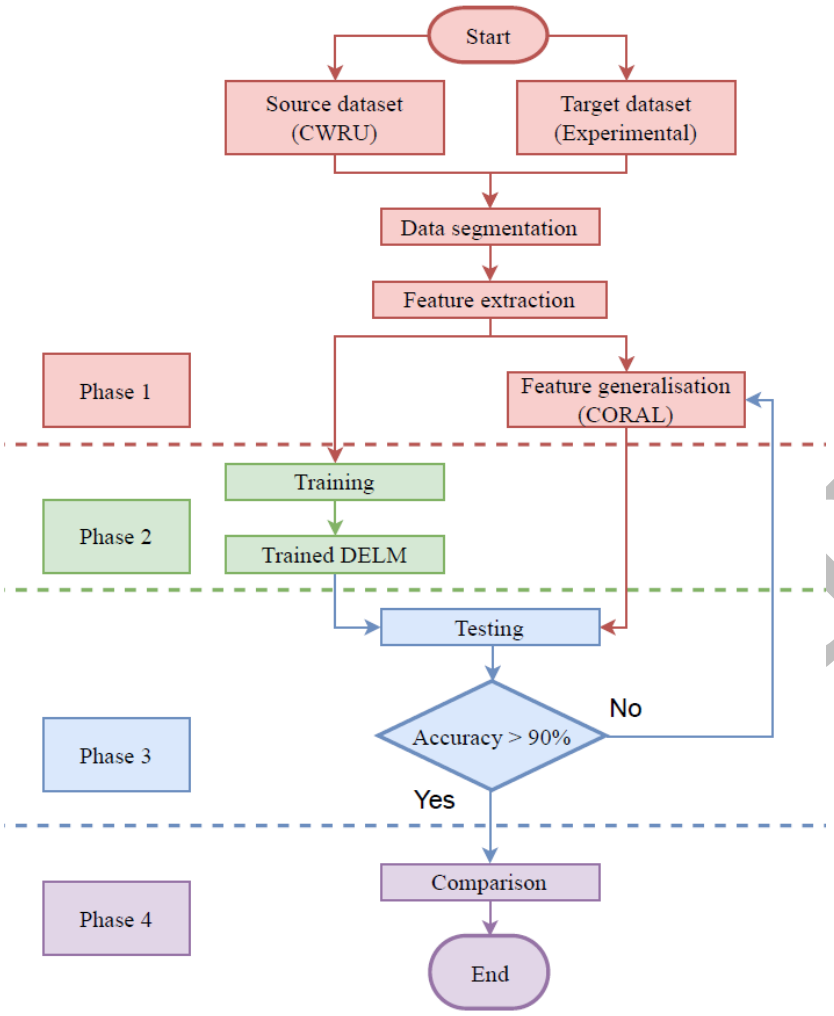

Figure 3. Project workflow

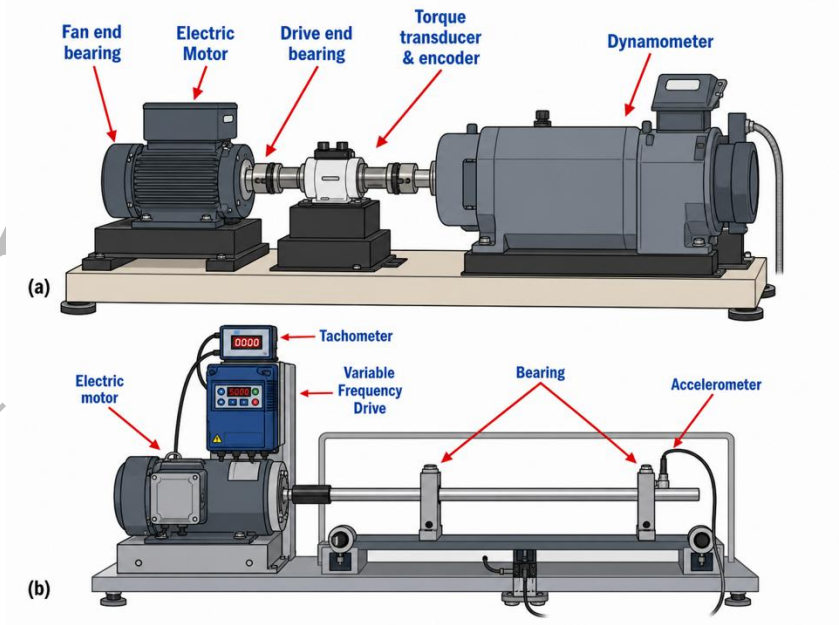

Figure 4. (a) CWRU test rig and (b) experimental test rig

Although frequency- and time-frequency-domain methods, such as the Fast Fourier Transform (FFT) and Continuous Wavelet Transform (CWT), have been widely adopted in fault diagnosis frameworks, they require additional preprocessing steps, thereby increasing computational complexity. Additionally, FFT is known to be sensitive to varying operating conditions and environments. To ensure simplicity, this study used only time-domain statistical features. Future studies may explore the potential performance improvement afforded by frequency- and time-frequency-domain methods.

Table 1. Sampled data configuration

Datasets	Condition	No of samples	Data per samples
CWRU	Healthy	100	400
	Ball fault	100	400
	Inner fault	100	400
	Outer fault	100	400
Experimental	Healthy	100	853
	Ball fault	100	853
	Inner fault	100	853
	Outer fault	100	853

Table 2. Time-domain features extracted from the samples

Features	Formula	Features	Formula
Mean	$\frac{1}{N}\sum_{i=1}^{N} x_i$	Peak-to-peak	$\max(x) - \min(x)$
Standard deviation	$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$	Margin	$(\frac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i \rvert})^2$
Root mean square (RMS)	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2}$	Energy	$\sum_{i=1}^{N} {x_i}^2$
Skewness	$\frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^3$	Log detector	$\frac{1}{N}\sum_{i=1}^{N}\log(\lvert x_i \rvert + \varepsilon)$
Kurtosis	$\frac{1}{N}\sum_{i=1}^{N}(\frac{x_i - \mu}{\sigma})^4$	Mean absolute difference	$\frac{1}{N-1}\sum_{i=1}^{N-1}\lvert x_{i+1} - x_i \rvert$

The next step was feature generalisation, in which CORAL was applied to reduce distributional differences between the experimental dataset (target domain) and the CWRU dataset (source domain). CORAL computed domain shift for each fault in both datasets, followed by the alignment of extracted features from both domains based on the corresponding fault conditions. The aim of implementing class-specific generalisation was to ensure that the distinguishing characteristics of each class were preserved, thereby yielding more accurate generalised features. Following feature generalisation, the extracted CWRU features were used to train the DELM model, while the generalised features from the experimental dataset were used for testing. In this study, the procedure described above corresponded to a forward testing case. This data separation strategy was designed to more effectively evaluate the model's ability to learn from the source domain and generalise to unseen data. The implementation of DELM was based on a publicly available open-source code repository [37] associated with the work reported in [38]. Once the training was done, the model proceeded with the testing phase. During the testing phase, unseen, generalised features were passed through DELM to ensure consistent feature transformation. The encoded features were then classified by the model's last layer, the ELM classifier. The DELM's predicted outputs were compared with the true labels to assess accuracy. To assess overall and class-level performance, metrics such as precision, recall, and F1 score were computed to capture class balance and misclassification behaviour across different classes. In addition, confusion matrices were generated to further support the findings from the previously mentioned metrics, particularly to identify misclassifications between classes. This approach offered deeper insights into the model's diagnostic capability.

To further assess the robustness of the proposed framework, one additional step was added: reverse testing. In this case, the roles of CWRU and the experimental dataset were reversed: the experimental dataset was treated as the source domain for training. In contrast, the CWRU dataset became the target domain for testing. The rest of the methodology and model architecture remained the same. Reverse testing was conducted to further validate the model's generalisation capability under different domain configurations. In the final phase, the proposed CORAL-DELM framework was compared with recent cross-machine bearing fault diagnosis frameworks, including DCFN [17], DDATN [18], PIUDA [19], and MTSTLF [20]. The classification accuracies were used as benchmarks to assess the effectiveness and competitiveness of the proposed framework in cross-machine applications.

3. Results and Discussion

The results of this paper follow a similar structure in the methodology. Phase 1 began with data acquisition and sampling using fixed windows, as listed in Table 1. The raw signals and sampled signals are illustrated in Figure 5. All sampled data were used for feature extraction. A total of 10 time-domain features were extracted and normalised from each sample. After feature extraction, feature generalisation was executed. This was done by applying CORAL to features from the CWRU and experimental datasets. The aim was to generalise the experimental dataset (target domain) to the CWRU dataset (source domain). Figure 6 depicts feature trends before and after generalisation, with each 100-sample index corresponding to a fault class. CORAL successfully aligned the features from experimental to mimic CWRU features. Although there are minor fluctuations, the overall patterns are similar, indicating that the generalisation process was carried out properly. This alignment helped the model better understand and classify the target domain, as the feature distributions of the two domains became less distinguishable. Moving on to the second phase, the training phase was done solely using CWRU features. CWRU features served as the foundation of the DELM model and were used to formulate it for testing. The DELM architectures designed for this study comprised three hidden layers with 64, 32, and 16 neurons, respectively. This configuration was selected through empirical trial and error, informed by preliminary experiments, to strike a balance between simplicity and classification performance. The ReLU activation function was adopted in the model for its simplicity, fast convergence, and widespread use in fault diagnosis. Figure 7 illustrates the DELM architecture, and the parameters and hyperparameters used in the DELM model are listed in Table 3.

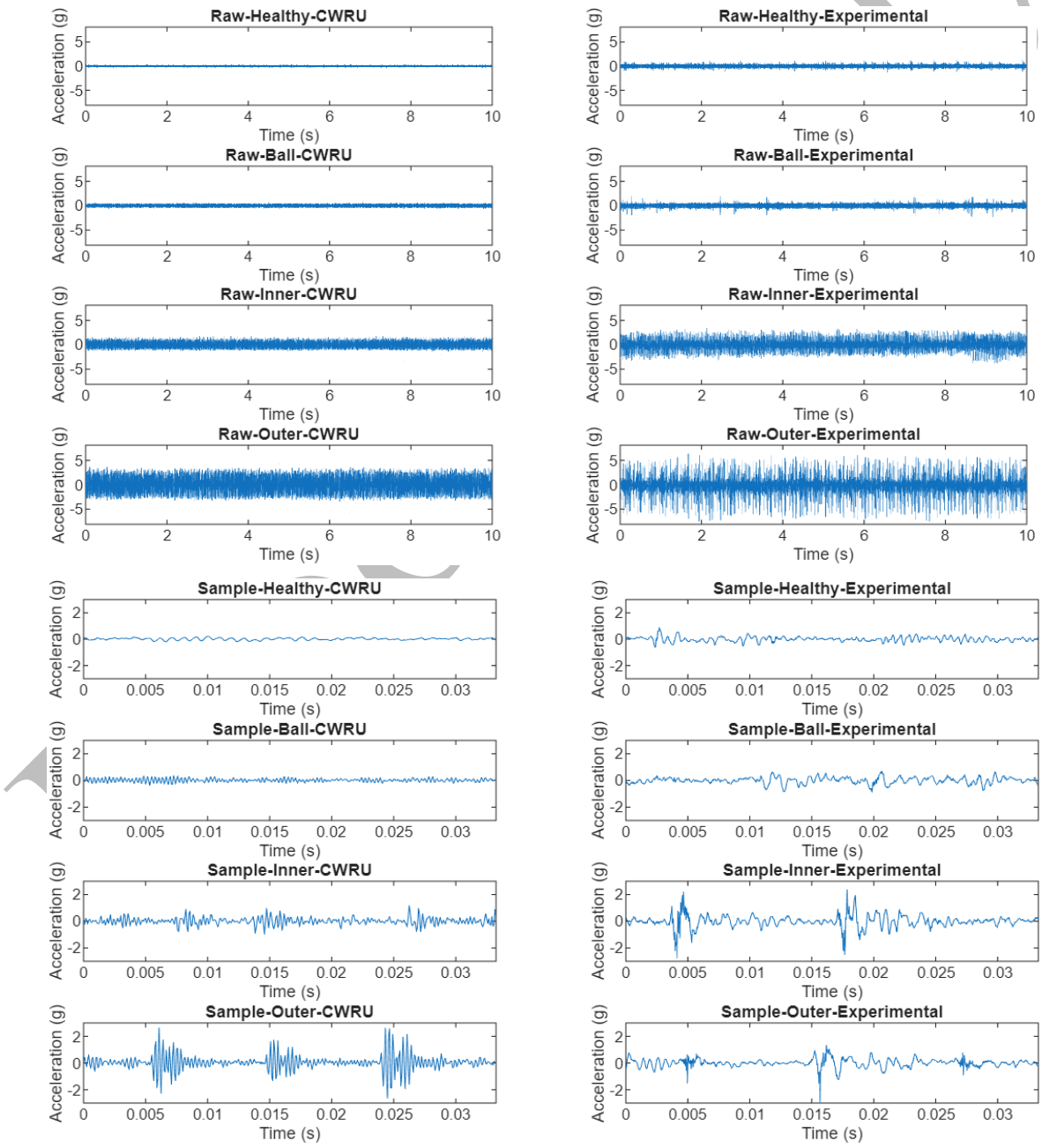

Figure 5. Raw and sampled vibration signals from CWRU and the experimental dataset

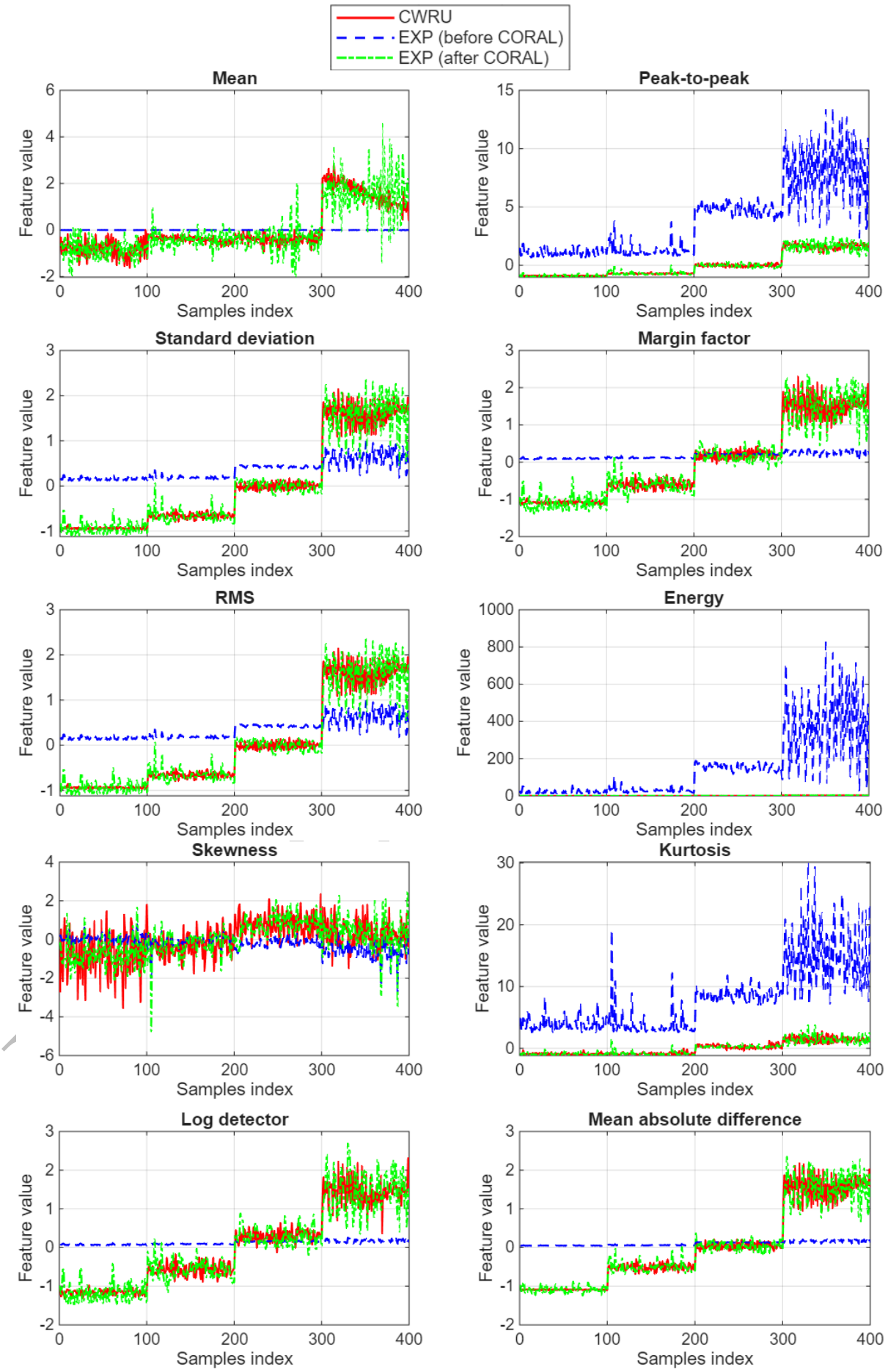

Figure 6. CORAL feature generalization after normalization

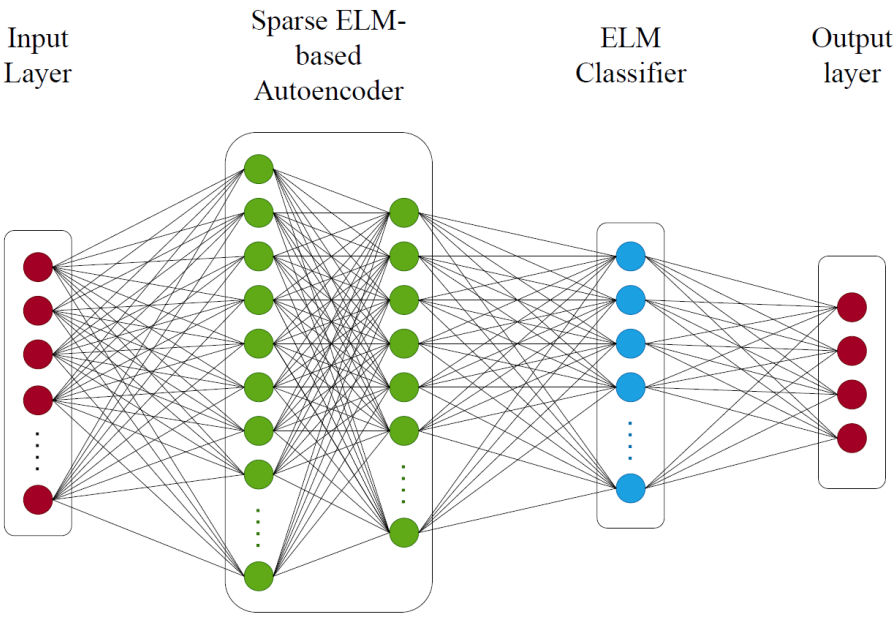

Figure 7. DELM architecture

Table 3. Summary of DELM architecture

Category	Parameter	Value/Range
ELM-Based Autoencoder	No. of layers	2
	No. of neurons	Layer 1: 64 Layer 2: 32
	Activation function	ReLU
	Epochs	100
	Regularization parameter	0.001
ELM Classifier	No. of neurons	16
	Regularization parameter	1
	Activation function	ReLU
Training Settings	Training dataset	CWRU (400 samples)
	Testing dataset	Experimental (400 samples)

Once the model was trained, it proceeded to the third phase: testing. In cross-machine fault diagnosis, testing performance was prioritised, as it defined the model's ability to generalise to unseen data across different machines. The testing phase involved utilising the trained DELM model to classify new, unseen features. These unseen features were the generalised features obtained in Phase 1. The result obtained from the proposed CORAL-DELM framework is summarised in Figure 8, followed by Tables 4 and 5. Figure 8 presents the accuracy distribution for both testing directions across 30 independent runs. As can be observed, the compact interquartile ranges for both testing directions, along with the low overall range, suggest low variability. In addition, Levene's test was conducted to assess whether the variances differed between the two tests, yielding a W statistic and a p-value that indicated statistical significance. Since $p > 0.05$, there was no significant difference in variances. Similarly, the small value of W proved the variances were nearly identical and not statistically different. Overall, the findings from Levene's test suggested that the proposed framework exhibited stable performance across 30 independent runs. However, a difference in accuracy remained between the two testing directions, with reverse testing exhibiting lower accuracy than forward testing. This finding can be explained by the feature distributions shown in Figure 6. As can be seen, the CWRU features demonstrated a clearer, well-separated pattern across all four classes, particularly in the mean, standard deviation, RMS, and other measures. The distinct, step-like distributions between classes suggested that CWRU features carried stronger class-separability information. Thus, it provides a well-structured feature space for the DELM to learn from. In contrast, the experimental features exhibited compressed, less-separated distributions, particularly for features such as the mean and margin factor. The class boundaries in the experimental feature distributions were less distinct. Following CORAL, the experimental features were shifted closer to the CWRU distribution, as reflected by the green dash-dotted line aligning more closely with the red line in Figure 6. In forward testing, since the DELM was trained on well-separated CWRU features, the model developed strong decision boundaries. This allowed the model to generalise effectively to the aligned experimental features. Conversely, in reverse testing, the DELM was trained on experimental features that exhibited weaker class separability.

Even after applying CORAL, the model struggled to learn effective decision boundaries. This explained why reverse testing typically yielded lower accuracy compared to forward testing.

Table 4. Overall performance of the proposed framework

Testing Direction	Metric	Value/Range
Forward Testing (CWRU → Experimental)	Accuracy (%)	92.95 ± 2.19
	Precision (%)	93.11 ± 2.11
	Recall (%)	92.95 ± 2.19
	F1-Score (%)	92.93 ± 2.21
	Runtime (s)	0.36 ± 0.01
Reverse Testing (Experimental → CWRU)	Accuracy (%)	89.58 ± 2.26
	Precision (%)	89.98 ± 2.15
	Recall (%)	89.58 ± 2.26
	F1-Score (%)	89.53 ± 2.29
	Runtime (s)	0.37 ± 0.04

Table 5. Class-level performance of the proposed framework

Testing Direction	Class	Precision (%)	Recall (%)	F1-Score (%)
Forward Testing (CWRU → Experimental)	Healthy	90.71 ± 4.76	91.30 ± 3.71	90.91 ± 3.11
	Ball fault	88.51 ± 4.48	86.80 ± 6.32	87.47 ± 4.00
	Inner-race fault	94.31 ± 2.81	96.97 ± 4.21	95.54 ± 2.47
	Outer-race fault	99.15 ± 1.30	96.63 ± 3.34	97.85 ± 1.98
Reverse Testing (Experimental → CWRU)	Healthy	81.92 ± 6.02	84.57 ± 7.82	82.93 ± 4.88
	Ball fault	84.42 ± 6.62	80.63 ± 8.25	82.12 ± 5.24
	Inner-race fault	94.44 ± 3.93	99.30 ± 1.34	96.76 ± 2.11
	Outer-race fault	99.14 ± 1.83	93.83 ± 4.54	96.34 ± 2.39

In Tables 4 and 5, the results are shown as the mean ± standard deviation across 30 independent runs to demonstrate the model's stability in classifying all four classes. For forward testing, the proposed framework achieved an average accuracy of 92.95%, with macro-averaged precision, recall, and F1 score of 93.11%, 92.95%, and 92.93%, respectively. The result demonstrated the model's ability to maintain balanced classification performance and shows that it did not bias towards any of the included classes. Apart from that, the results showed small standard deviations, indicating high consistency and reproducibility under random initialisation, which is highly recommended for cross-machine applications. Even with weaker class separability, the proposed framework achieved reliable performance in reverse testing, averaging 89.58% accuracy, with corresponding macro-averaged precision, recall, and F1 score of 89.98%, 89.58%, and 89.53%, respectively. Despite a slight reduction in performance relative to forward testing, the results remained stable, suggesting the proposed framework's ability to maintain reliable performance even when the domain roles were reversed. Furthermore, the similar runtimes of the two tests further demonstrate the framework's efficiency in cross-machine fault diagnosis.

Table 5 presents class-level performance for both testing directions, including precision, recall, and F1 score. Precision quantifies the proportion of true positive identifications relative to the total number of positive classifications. In a diagnostic context, it represents the probability that a detected anomaly or fault condition truly exists. Recall, often referred to as sensitivity or the true positive rate, quantifies the proportion of actual fault instances that the diagnostic model correctly identifies. It reflects the model's efficacy in capturing the full set of positive-class samples in the ground-truth data. The F1-score is defined as the harmonic mean of precision and recall, providing a single scalar metric that balances diagnostic reliability with detection coverage. It is particularly advantageous for evaluating models trained on imbalanced datasets, as it penalises significant disparities between the two constituent metrics. Consequently, a high F1-score serves as a rigorous indicator of a model's overall classification robustness. In forward testing (see Table 5), inner-race and outer-race faults achieved high scores across all three metrics, indicating that the selected time-domain statistical features effectively captured their vibration characteristics and that CORAL-based domain adaptation was effective. However, the ball fault class exhibited relatively lower class-level performance metrics, with averages of 88.51% precision, 86.80% recall, and 87.47% F1-score. This suggests a higher likelihood that ball faults are misclassified. This outcome might stem from the overlapping statistical feature distributions between the ball fault and healthy conditions, as both classes exhibited lower class-level performance metrics. This behaviour was consistent with the confusion matrices shown in Figure 9, where most misclassifications occurred between ball fault and healthy conditions. In contrast, inner- and outer-race faults remained well classified. The similarities between the two were further demonstrated by the raw and sampled vibration signals shown in Figure 5, which exhibit comparable intensity and characteristics. This factor reduced separability, making it more challenging for the model to classify effectively. In reverse testing, similar trends

were observed in Table 5: both inner- and outer-race faults were well classified, while healthy and ball faults showed slightly reduced separability. Overall, the proposed framework has demonstrated promising generalisation and classification capabilities, with balanced performance across all four classes while maintaining low computational complexity.

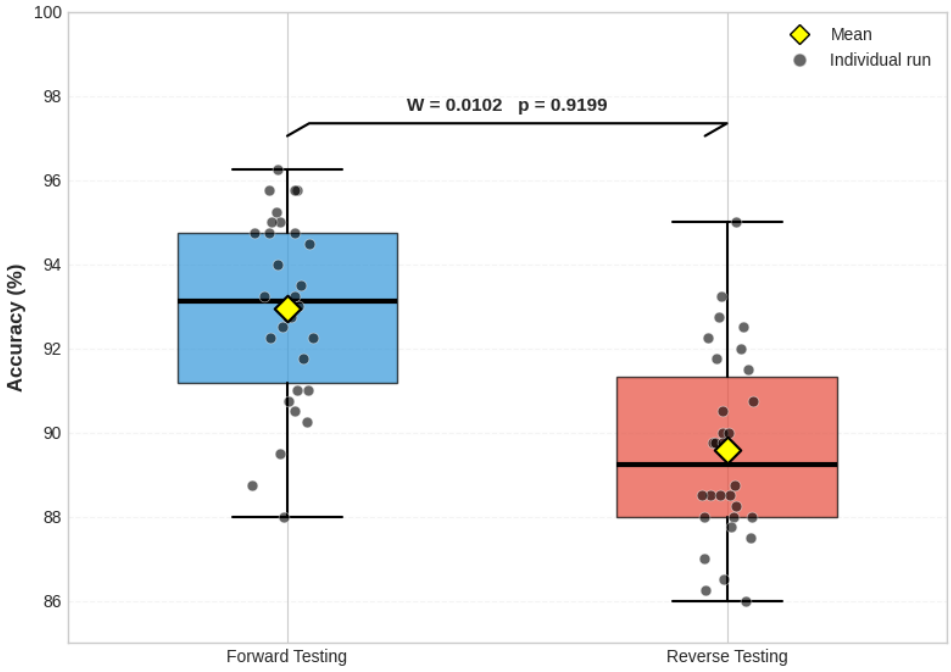

Figure 8. Model accuracy across 30 independent runs

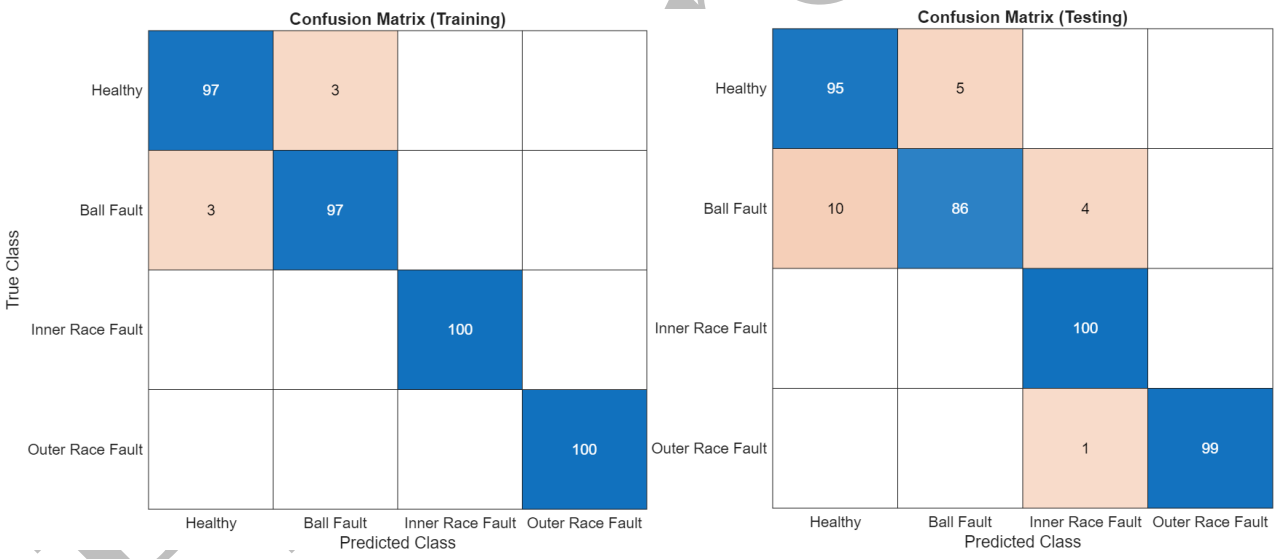

Figure 9. Confusion matrix for training and testing

Following the testing phase, the performance of CORAL-DELM was further evaluated by comparing both classification accuracy and computational runtime to assess its efficiency. The result is shown in Figure 10. In this study, efficiency is defined as achieving better performance while maintaining lower computational cost. As shown in Figure 10(a), CORAL-DELM improved diagnostic performance over its baseline model, ELM, achieving the highest accuracy, as listed in Table 6. On the other hand, Figure 10(b) shows that the runtime of CORAL-DELM is highly comparable to that of other models, with only small differences. The results in both figures demonstrate CORAL-DELM's ability to achieve a favourable trade-off between accuracy and runtime. CORAL-DELM improved diagnostic performance without introducing a significant computational burden. This advantage is desirable for practical cross-machine fault diagnosis applications, where performance must be balanced against computational constraints. Lastly, the proposed CORAL-DELM framework was compared with four state-of-the-art models as mentioned in Section 2. The accuracy comparison is demonstrated in Figure 11. It is important to note that the results shown are obtained under different datasets, experimental settings, and modelling assumptions. Therefore, the comparison is not intended for direct benchmarking but to provide valuable insights into the competitiveness and potential of the proposed framework in cross-machine applications. CORAL-DELM achieved a competitive level of accuracy compared to existing models. This

comparable diagnostic performance suggests that CORAL-DELM can achieve high classification accuracy relative to more advanced, complex model architectures, enabling future studies with more datasets under various operating conditions.

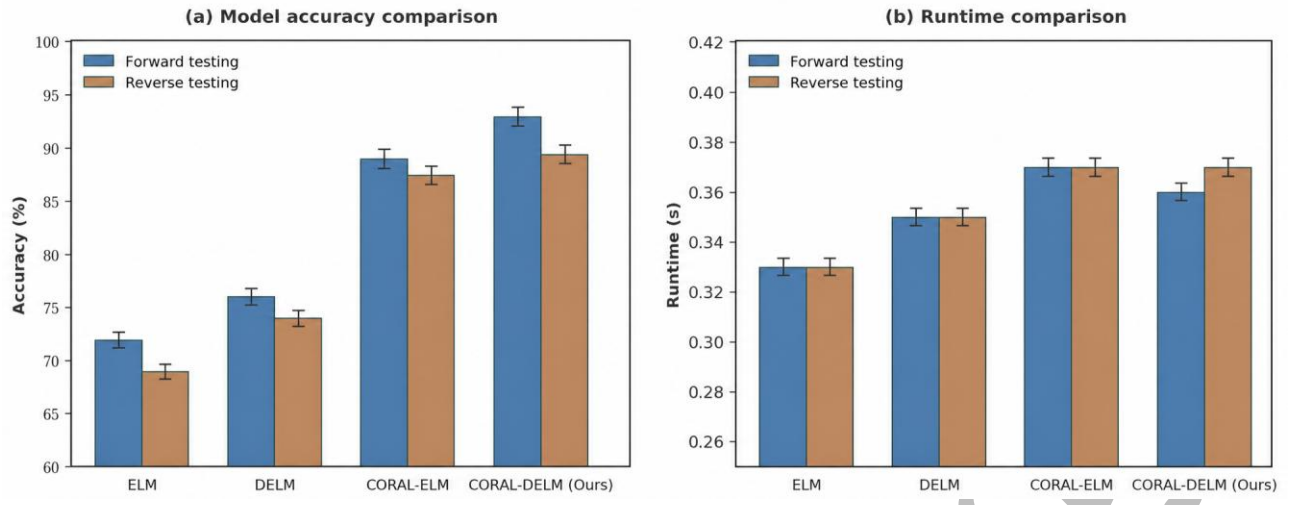

Figure 10. Performance comparison under forward and reverse testing conditions: (a) model accuracy comparison and (b) runtime comparison

Table 6. Accuracy and runtime performance under forward and reverse testing conditions

Model	Forward accuracy (%)	Reverse accuracy (%)	Average accuracy (%)	Forward runtime (s)	Reverse runtime (s)	Average runtime (s)
ELM	71.96 ± 2.73	68.75 ± 3.15	70.36	0.33 ± 0.01	0.33 ± 0.01	0.33
DELM	75.94 ± 2.14	74.08 ± 5.14	75.01	0.35 ± 0.05	0.35 ± 0.04	0.35
CORAL-ELM	88.89 ± 2.76	87.40 ± 3.66	88.15	0.37 ± 0.03	0.37 ± 0.03	0.37
CORAL-DELM (Ours)	92.95 ± 2.19	89.58 ± 2.26	91.27	0.36 ± 0.01	0.37 ± 0.04	0.37

3.1 Limitations and future work

Despite having a promising performance, several limitations in the proposed framework should be addressed. First, this study uses only two datasets: the CWRU bearing dataset and the experimental dataset. Although the evaluation is sufficient to demonstrate cross-machine reliability, the effectiveness of CORAL-DELM in broader industrial settings involving more than two datasets remains limited. The broader scenarios may include multiple machines, sensing configurations, and different operating conditions. Second, the CORAL algorithms used in this study may align the two domains based on second-order statistics, but may be insufficient when domain differences are significant and nonlinear. Third, all datasets used and the experiments conducted in this study were under fixed operating conditions. The effects of various operating conditions, such as motor speed and load, were not accounted for. All the factors mentioned above may significantly affect vibration signals, thereby reducing the separability of the extracted features. Future work will incorporate additional datasets, explore more advanced domain adaptation approaches, and examine different operating conditions to further assess the robustness of the proposed framework.

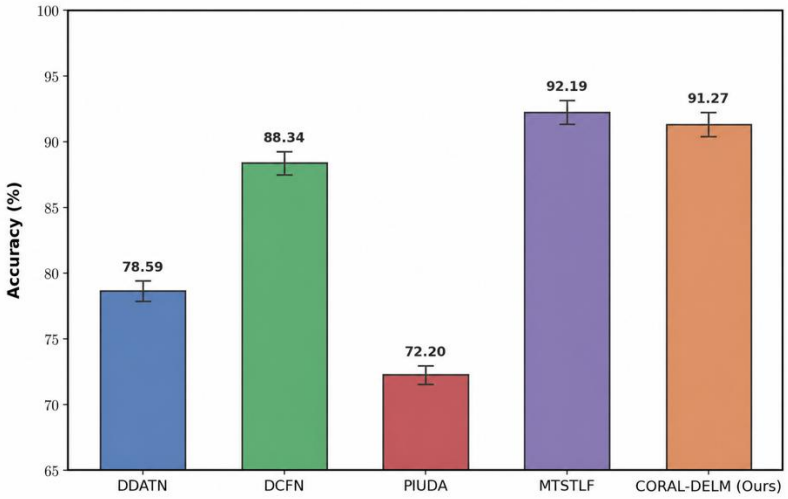

Figure 11. Comparison with state-of-the-art models

4. Conclusions

This paper presents a simple and efficient CORAL-DELM framework as an alternative learning model for cross-machine bearing fault diagnosis, aiming to improve generalisation capability while prioritising simplicity and efficiency. This study utilised CORAL feature alignment to minimise distributional differences between CWRU and experimental datasets. Time-domain statistical features were extracted from the raw vibration signals and used as input to DELM for classification. The proposed framework demonstrated competitive performance, achieving classification accuracy of 92.95% ± 2.19% and 89.58% ± 2.26% for forward and reverse testing, respectively. This shows the stability of CORAL-DELM in classifying four bearing fault conditions across domains. The model's efficiency has also been demonstrated, with average runtimes of 0.36 ± 0.02s and 0.33 ± 0.08s for forward and reverse testing, respectively, indicating its potential for practical deployment. These findings suggest that an effective cross-machine fault diagnosis system can be achieved without deep networks or complex transfer learning architectures. From an industrial standpoint, CORAL-DELM enables engineers to perform effective fault diagnosis across multiple machines without extensive data collection or frequent model retraining. The proposed framework's diagnostic performance demonstrates its applicability to real-time applications, helping prevent machine breakdowns and reduce maintenance costs.

Acknowledgements

The authors thank the Ministry of Higher Education under the Fundamental Research Grant Scheme for financial support and Universiti Teknologi Malaysia for laboratory facilities and financial support.

Funding

This work was supported by the Ministry of Higher Education under the Fundamental Research Grant Scheme (Grant Number: FRGS/1/2023/TK10/UTM/02/12) and Universiti Teknologi Malaysia under the UTM Kecerdasan Buatan grant (Grant Number: Q.J130000.5724.44H57).

Declaration of Competing Interest

The author declares no conflicts of interest.

CRediT Authorship Contribution Statement

Muhammad Amir Zikry Harun: Conceptualisation; Methodology; Writing-draft
Muhammad Firdaus Isham: Methodology; Analysis; Validation; Writing-revise & review
Mohd Syahril Ramadhan Saufi: Methodology; Analysis; Software; Writing-revise & review
Wan Aliff Abdul Saad: Methodology; Writing-revise & review
Muhammad Danial Abu Hasan: Methodology; Investigation; Writing-revise & review

Availability of Data and Materials

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Statement

This study did not involve human participants or animal subjects. Ethical approval was therefore not required for this research.

Generative Artificial Intelligence Declarations

The authors claim that artificially intelligent-assisted technologies, such as generative AI, were not used to generate content, ideas, or theories. We have just utilised AI to enhance readability and refine the language. This was used with extreme human control and oversight. The authors take full responsibility for the review and approval of the content.

References

[1] M. Romanssini, P. C. C. de Aguirre, L. Compassi-Severo, and A. G. Girardi, "A review on vibration monitoring techniques for predictive maintenance of rotating machinery," *Eng*, vol. 4, no. 3, pp. 1797–1817, 2023. https://doi.org/10.3390/eng4030102

[2] S. K. Yang, C. M. Chen, and H. L. Chang, "A combined preventive maintenance strategy for bearings to accomplish the failure prevention of rotating equipment," *Journal of Failure Analysis and Prevention*, vol. 22, no. 4, pp. 1457–1467, 2022. https://doi.org/10.1007/s11668-022-01415-8

[3] L. Pinciroli, P. Baraldi, and E. Zio, "Maintenance optimization in industry 4.0," *Reliability Engineering & System Safety*, vol. 234, p. 109204, 2023. https://doi.org/10.1016/j.ress.2023.109204

[4] J. Cen, Z. Yang, X. Liu, J. Xiong, and H. Chen, "A review of data-driven machinery fault diagnosis using machine learning algorithms," *Journal of Vibration Engineering & Technologies*, vol. 10, no. 7, pp. 2481–2507, 2022. https://doi.org/10.1007/s42417-022-00498-9

[5] I. U. Hassan, K. Panduru, and J. Walsh, "An In-depth study of vibration sensors for condition monitoring," *Sensors*, vol. 24, no. 3, 2024. https://doi.org/10.3390/s24030740

[6] Y. I. Alzoubi, A. Mishra, and A. E. Topcu, *Research trends in deep learning and machine learning for cloud computing security*, vol. 57, no. 5. Springer Netherlands, 2024. https://doi.org/10.1007/s10462-024-10776-5

[7] F. Zhang, X. Wang, and X. Zhang, "Applications of deep learning method of artificial intelligence in education," *Education and Information Technologies*, vol. 30, no. 2, pp. 1563–1587, 2025. https://doi.org/10.1007/s10639-024-12883-w

[8] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 186, no. October 2022, p. 109884, 2023. https://doi.org/10.1016/j.ymssp.2022.109884

[9] Z. Zhu Y. Lei, G. Qi, Y. Chai, N. Mazur, Y. An et al., "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery," *Measurement*, vol. 206, p. 112346, 2023. https://doi.org/10.1016/j.measurement.2022.112346

[10] Y. P. Zhao and Y. B. Chen, "Extreme learning machine based transfer learning for aero engine fault diagnosis," *Aerospace Science and Technology*, vol. 121, p. 107311, 2022. https://doi.org/10.1016/j.ast.2021.107311

[11] P. Dutta, D. Gupta, and J. Maurya, "Depression detection using extreme learning machine," *Proc. - n2024 4th International Conference on Pervasive Computing and Social Networking*, pp. 42–47, 2024. https://doi.org/10.1109/ICPCSN62568.2024.00015

[12] S. D. S. Dass, G. Krishnasamy, R. Paramesran, and R. C.-W. Phan, "Hessian unsupervised extreme learning machine," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 5, pp. 2013–2022, 2024. https://doi.org/10.1007/s13042-023-02012-3

[13] X. Han, S. Ma, Z. Shi, G. An, Z. Du and C. Zhao, "A novel power transformer fault diagnosis model based on harris-hawks-optimization algorithm optimized kernel extreme learning machine," *Journal of Electrical Engineering & Technology*, vol. 17, no. 3, pp. 1993–2001, 2022. https://doi.org/10.1007/s42835-022-01000-x

[14] N. Shan, X. Xu, X. Bao, and S. Qiu, "Fast Fault Diagnosis in Industrial Embedded Systems Based on Compressed Sensing and Deep Kernel Extreme Learning Machines," *Sensors*, vol. 22, no. 11, 2022. https://doi.org/10.3390/s22113997

[15] C. Zhao, E. Zio, and W. Shen, "Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study," *Reliability Engineering & System Safety*, vol. 245, p. 109964, 2024. https://doi.org/10.1016/j.ress.2024.109964

[16] Y. Shi, A. Deng, X. Ding, S. Zhang, S. Xu, and J. Li, "Multisource domain factorization network for cross-domain fault diagnosis of rotating machinery: An unsupervised multisource domain adaptation method," *Mechanical Systems and Signal Processing*, vol. 164, p. 108219, 2022. https://doi.org/10.1016/j.ymssp.2021.108219

[17] S. Jia, Y. Li, X. Wang, D. Sun, and Z. Deng, "Deep causal factorization network: A novel domain generalization method for cross-machine bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 192, p. 110228, 2023. https://doi.org/10.1016/j.ymssp.2023.110228

[18] Y. Zhou, Y. Dong, H. Zhou, and G. Tang, "Deep dynamic adaptive transfer network for rolling bearing fault diagnosis with considering cross-machine instance," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021. https://doi.org/10.1109/TIM.2021.3112800

[19] N. Jia, W. Huang, C. Ding, J. Wang, and Z. Zhu, "Physics-informed unsupervised domain adaptation framework for cross-machine bearing fault diagnosis," *Advanced Engineering Informatics*, vol. 62, no. PC, p. 102774, 2024. https://doi.org/10.1016/j.aei.2024.102774

[20] L. Zhao, Y. He, D. Dai, X. Wang, H. Bai, and W. Huang, "A novel multi-task self-supervised transfer learning framework for cross-machine rolling bearing fault diagnosis," *Electronics*, vol. 13, no. 23, p. 4622, 2024. https://doi.org/10.3390/electronics13234622

[21] J. Hendriks, P. Dumond, and D. A. Knox, "Towards better benchmarking using the CWRU bearing fault dataset," *Mechanical Systems and Signal Processing*, vol. 169, p. 108732, 2022. https://doi.org/10.1016/j.ymssp.2021.108732

[22] C. Wang, Y. Sun, and X. Wang, "Image deep learning in fault diagnosis of mechanical equipment," *Journal of Intelligent Manufacturing*, vol. 35, no. 6, pp. 2475–2515, 2024. https://doi.org/10.1007/s10845-023-02176-3

[23] A. Ghorbel, S. Eddai, B. Limam, N. Feki, and M. Haddar, "Bearing fault diagnosis based on artificial intelligence methods: machine learning and deep learning," *Arabian Journal for Science and Engineering*, vol. 50, no. 17, pp. 13605–13622, 2025. https://doi.org/10.1007/s13369-024-09488-3

[24] P. Ding, Y. Xu, P. Qin, and X. M. Sun, "A novel deep learning approach for intelligent bearing fault diagnosis under extremely small samples," *Applied Intelligence*, vol. 54, no. 7, pp. 5306–5316, 2024. https://doi.org/10.1007/s10489-024-05429-7

[25] E. Gürsoy and Y. Kaya, "An overview of deep learning techniques for COVID-19 detection: methods, challenges, and future works," *Multimedia Systems*, vol. 29, no. 3, pp. 1603–1627, 2023. https://doi.org/10.1007/s00530-023-01083-0

[26] B. A. Tama, M. Vania, S. Lee, and S. Lim, *Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals*, vol. 56, no. 5. Springer Netherlands, 2023. https://doi.org/10.1007/s10462-022-10293-3

[27] B. Peng, H. Xia, X. Lv, M. Annor-Nyarko, S. Zhu, Y. Liu et al., "An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network," *Applied Intelligence*, vol. 52, no. 3, pp. 3051–3065, 2022. https://doi.org/10.1007/s10489-021-02555-4

[28] J. Wang, S. Lu, S. H. Wang, and Y. D. Zhang, "A review on extreme learning machine," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 41611–41660, 2022. https://doi.org/10.1007/s11042-021-11007-7

[29] M. F. Isham, M. S. Leong, M. H. Lim, and Z. A. Ahmad, "Intelligent wind turbine gearbox diagnosis using VMDEA and ELM," *Wind Energy*, vol. 22, no. 6, pp. 813–833, 2019. https://doi.org/10.1002/we.2323
[30] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical Systems and Signal Processing*, vol. 133, p. 106272, 2019. https://doi.org/10.1016/j.ymssp.2019.106272
[31] L. Shi, S. Su, W. Wang, S. Gao, and C. Chu, "Bearing fault diagnosis method based on deep learning and health state division," *Applied Sciences*, vol. 13, no. 13, 2023. https://doi.org/10.3390/app13137424
[32] B. Zhang, H. Dong, H. A. A. M. Qaid, and Y. Wang, "Deep domain adaptation with correlation alignment and supervised contrastive learning for intelligent fault diagnosis in bearings and gears of rotating machinery," *Actuators*, vol. 13, no. 3, p. 93, 2024. https://doi.org/10.3390/act13030093
[33] X. He, "Quantum correlation alignment for unsupervised domain adaptation," *Physical Review A*, vol. 102, no. 3, pp. 1–11, 2020. https://doi.org/10.1103/PhysRevA.102.032410
[34] B. Sun, J. Feng, and K. Saenko, "Correlation alignment for unsupervised domain adaptation," *Advances in Computer Vision and Pattern Recognition*, pp. 153–171, 2017. https://doi.org/10.1007/978-3-319-58347-1_8
[35] P. Singhal, R. Walambe, S. Ramanna, and K. Kotecha, "Domain adaptation: challenges, methods, datasets, and applications," *IEEE Access*, vol. 11, pp. 6973–7020, 2023. https://doi.org/10.1109/ACCESS.2023.3237025
[36] Q. Qian, Y. Qin, Y. Wang, and F. Liu, "A new deep transfer learning network based on convolutional auto-encoder for mechanical fault diagnosis," *Measurement*, vol. 178, p. 109352, 2021. https://doi.org/10.1016/j.measurement.2021.109352
[37] A. Alitaleshi, "Deep-Extreme-Learning-Machine." Accessed: Jun. 09, 2026. [Online]. Available: https://github.com/atefeta/Deep-Extreme-Learning-Machine
[38] A. Alitaleshi, H. Jazayeriy, and J. Kazemitabar, "Affinity propagation clustering-aided two-label hierarchical extreme learning machine for Wi-Fi fingerprinting-based indoor positioning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 6, pp. 3303–3317, 2022. https://doi.org/10.1007/s12652-022-03777-1