

Integrating Spatial and Temporal Features for Bearing Fault Diagnosis: A Cross-Domain Analysis Using Vibration and Acoustic Data

Mert Sehri¹, Niousha Khalilian¹, Francisco de Assis Boldt², Michel Bouchard³, Patrick Dumond¹

¹*University of Ottawa, Ottawa, ON, K1N 6N5, Canada*

msehr006@uottawa.ca

nkhal007@uottawa.ca

pdumond@uottawa.ca

²*Instituto Federal do Espírito Santo, Vitoria, ES, 29173-087, Brazil*

franciscoa@ifes.edu.br

³*General Bearing Service Inc., 490 Kent Street, Ottawa, ON, K2P 2B7, Canada*

michelb@gbs.ca

ABSTRACT

Bearing failures cause machinery breakdowns, resulting in financial losses due to production downtimes. To address this, accurate bearing condition monitoring is essential. This paper introduces a cross-domain approach to fault diagnosis using a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) models, applied to the Case Western Reserve University (CWRU) dataset and the University of Ottawa Rolling-element Dataset- Vibration and Acoustic Faults under Constant Load and Speed conditions (UORED-VAFCLS), which contain both artificial and naturally developed bearing faults. The proposed experimental framework assesses the estimators, training and testing them with raw time-domain data from both acoustic and accelerometer signals, enhancing fault detection across various operating conditions. Results demonstrate that the CNN-LSTM model, when combined with statistical preprocessing, outperforms advanced models in both performance, computational time, and stability, particularly when fusing data from multiple sources. This approach shows promise for practical implementations in industrial predictive maintenance, offering a more reliable solution for reducing downtime and improving operational efficiency. Future work will focus on further optimization of the model and minimizing the data required for effective condition monitoring.

Mert Sehri et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.36001/IJPHM.2025.v16i2.4469>

1. INTRODUCTION

In industrial applications, 50% to 60% of machinery failures are caused by bearings (Jha & Swami, 2021; Sehri, Dumond, & Bouchard, 2023a). These failures set companies back millions of dollars every year in production maintenance downtime, and in some cases lead to catastrophic failures (Sehri et al., 2023a). This creates a necessity to explore whether machine learning (ML) algorithms can be used to detect bearing faults, helping to minimize maintenance downtime. Bearing condition monitoring includes classification of signals obtained from healthy, ball, cage, inner-race, and outer-race fault states. With the evolution of artificial intelligence (AI), condition monitoring is made easier by decreasing the need for expertise. Specifically, with deep learning (DL) algorithms, ML researchers can now train and test models using signals obtained from sensors such as accelerometers and microphones to validate a machine's health state against different signals with similar classifications (Bayouhdh, 2024).

In ML, the domain refers to the raw sensor files chosen from a bearing dataset to create suitable testing conditions for fault diagnosis. Understanding the domain is crucial for effective feature extraction, model design, and interpreting the results of the ML architecture (Bayouhdh, 2024). A domain refers to a combination of bearings operating under similar load or speed conditions. In other cases, a domain could represent bearings with similar fault sizes or types operating under different conditions. On the other hand, cross-domain analysis in ML refers to a ML algorithm trained on data from one domain that is adapted to work effectively in another

domain (Layeghy & Portmann, 2023). The purpose is to use the source domain knowledge to improve performance accuracy (bearing fault detection) in the target domain (Yu, Karimi, Shi, Peng, & Zhao, 2024).

Many researchers have focused on training ML algorithms for bearing fault classification by using the Case Western Reserve University (CWRU) dataset as a benchmark (Hendriks, Dumond, & Knox, 2022). The CWRU dataset provides rolling element bearing data containing artificially created “seeded” faults. Moreover, this dataset has been criticized as having similar load conditions (motor speeds), making it easy to obtain high validation accuracy across different load condition domains. However, if domains are defined by fault severities instead, performance drops significantly (Hendriks et al., 2022). In fact, cross-domain accuracies are usually above 95% when using CWRU bearing data divided by load conditions (speeds) (Zixian Li et al., 2024), while splitting the data based on fault sizes gives 50-60% accuracy (Hendriks et al., 2022; Rauber, da Silva Loca, Boldt, Rodrigues, & Varejão, 2021). This is because in the first case, the same bearing is used to both train and test the ML algorithm, whereas in the latter case, different bearings with unique fault signatures are used. Furthermore, when a natural fault dataset, such as the University of Ottawa Rolling-element Dataset – Vibration and Acoustic Faults under Constant Load and Speed conditions (UORED-VAFCLS), is used with raw time domain data instead, even after preprocessing with statistical methods and applying an ML algorithm for cross-domain fault detection, accuracies are expected to be even worse because natural faults are more complex than controlled (seeded) faults. More advanced ML methods or more sophisticated methods of data processing must be applied to reach higher accuracy.

DL has been increasingly adopted in cross-domain classification tasks for bearing fault detection due to its ability to automatically learn and generalize features from raw sensor data (Chen et al., 2020). In practical industrial applications, variations in operational conditions, such as changes in load, speed, or fault severity (amplitude changes), complicate fault diagnosis (Lundström & O’Nils, 2023). Cross-domain classification methods attempt to address this challenge by utilizing ML algorithms to adapt knowledge from a source domain (e.g., one set of operational conditions) to a target domain (e.g., different operational conditions) (Yin, Chen, Luo, & Deng, 2023). Specifically, DL networks like convolutional neural networks (CNNs) are able to capture spatial features from vibration signals (Wang, Liu, Peng, Yang, & Qin, 2022), while long short-term memory (LSTM) networks are effective in modeling temporal dependencies inherent in sensor data (Wan, Guo, Yin, Liang, & Lin, 2020). By combining these networks, cross-domain classification becomes more robust, allowing the algorithms to handle variations in input data more effectively (Xu, Yu, Chen, & Lin, 2024). Moreover, the preprocessing of raw signals using statistical techniques can reduce noise and highlight important patterns, further enhancing the

adaptability and accuracy of these models in diverse domains (Meng, Zhan, Li, & Pan, 2018). This approach is particularly valuable for industries seeking scalable and reliable fault diagnosis solutions without the need for extensive domain-specific expertise. By automating feature learning, industry can implement AI-driven solutions without requiring specialists to manually analyze sensor data. As such, the integration of DL methods into cross-domain classification represents a promising direction for advancing predictive maintenance in industrial settings.

The literature base for prognostic health management (PHM) and industrial AI can also be broadened by considering recent works that provide both methodological and conceptual advancements in the field. For example, Su and Lee provide detailed insights into leveraging publicly available PHM datasets for diagnostic and prognostic modeling (Su & Lee, 2024), while Lee and Su present a unified conceptual framework for integrating industrial AI into manufacturing and maintenance systems (Lee & Su, 2025). Including such perspectives strengthens the contextual foundation of this study, situating it more firmly within ongoing advancements in PHM and industrial AI research.

This paper explores the effects of combining an LSTM network, a type of recurrent neural network (RNN), with a convolutional neural network (CNN) to see how cross-domain performance can be improved for bearing fault detection, particularly in datasets with naturally occurring faults (e.g., the UORED-VAFCLS dataset). Additionally, the paper explores the benefits of using simple statistical methods for preprocessing data, showing that these techniques can enhance ML model accuracy. By applying this traditional approach to bearing fault detection, it is hoped that performance can remain similar to more advanced methods but at significantly lower computational costs. By leveraging the strengths of both spatial and temporal feature extraction, this hybrid approach allows for generalization across varying conditions, highlighting the effectiveness of traditional methods when paired with data-driven optimization.

This paper is organized as follows: Section 2 presents the theoretical foundation of cross-domain fault diagnosis and the integration of CNN-LSTM models for bearing condition monitoring. Section 3 introduces the methodology, detailing the proposed model architecture and preprocessing techniques. Section 4 focuses on experimental verification, displaying the results of applying the model to both the CWRU and the UORED-VAFCLS datasets with an emphasis on cross-domain accuracy and robustness. Finally, Section 5 concludes the paper by summarizing the contributions and discussing potential avenues for future work.

1.1. Motivation

The design choices in this study were guided by the need to balance model performance, generalization ability, and computational efficiency in cross-domain bearing fault

diagnosis. The CNN-LSTM architecture was selected over more recent architectures such as Transformers with temporal embedding and temporal convolutional networks because, while promising, these alternatives are computationally inefficient compared to the proposed method. They require greater training times, larger memory resources, and extensive hyperparameter tuning to achieve stable convergence (Sehri, Hua, Boldt, & Dumond, 2025). Given the already extensive experimental space explored in this work covering multiple preprocessing strategies, multimodal fusion, and strict domain separation, the priority was to validate a lightweight hybrid ML model before expanding to more resource-intensive approaches.

A similar logic applies to the multimodal fusion strategy, where simple feature concatenation was used to combine acoustic and vibration data. This approach helps to establish a clear performance baseline and isolate the benefits of multimodal integration without the added complexity of advanced fusion mechanisms such as attention-based or correlation-aware methods, which will be explored in later studies.

Furthermore, interpretability and explainability methods such as Shapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME) were not implemented in this work to maintain focus on the primary cross-domain classification objective. These tools will be incorporated in future research to better understand the contribution of different input features in making predictions and to improve integration in industrial applications.

2. THEORY

To build on the concepts introduced herein, this section delves into the theoretical underpinnings of the proposed CNN-LSTM architecture, detailing its structural innovation, adaptation to cross-domain challenges, and the methodologies used to enhance its performance. This section provides an understanding of the technical design choices made in this study and provides a direction for experimental validation.

The CNN-LSTM architecture integrates the strengths of CNNs and LSTM networks to effectively address cross-domain fault diagnosis challenges. This section provides a detailed theoretical basis for the choice and functionality of these components, emphasizing their role in handling both spatial and temporal patterns in fault data.

CNNs are widely recognized for their ability to extract spatial hierarchies of features through learnable filters and pooling operations. The mathematical foundation of CNNs relies on the convolution operation (Lecun, Bottou, Bengio, & Haffner, 1998; McCarter, 2023):

$$f \cdot \omega(x) = \int f(y)\omega(x-y)dy \quad (1)$$

where f and ω represent the input function and kernel function respectively, “ $\cdot$ ” is the convolution operation, x and y are real numbered variables. By utilizing CNNs at the input

stage, high-dimensional vibration or sensor data can be transformed into low-level representations, capturing critical spatial patterns such as root mean square (RMS) features in vibration signals. This step is crucial for domain-invariant feature extraction, as it reduces sensitivity to noise and domain variations.

To capture temporal dependencies inherent in fault diagnosis data (e.g., sequential vibration signals), LSTM networks are introduced. LSTMs are a type of RNN that are designed to overcome vanishing gradient problems through their specialized gating mechanism. Each LSTM cell operates as defined by the following equations (Hochreiter & Schmidhuber, 1997):

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$= f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

where i_t , f_t , and o_t are the input, forget, and output gates, respectively; c_t represents the cell state; and $\odot$ is the element-wise product. The input gate controls how much of the new input and the previous hidden state should contribute to the memory update. The forget gate determines what portion of the previous cell state should be retained. The output gate decides how much of the updated cell state is exposed to influence the next hidden state. By processing the sequential feature maps generated by the CNN layers, LSTM cells capture temporal correlations between faults over time, addressing data with dynamic variations and long-range dependencies.

The integration of CNN and LSTM components allows the architecture to adapt to both spatial and temporal variations present in cross-domain data. To further enhance domain generalization, domain-invariant feature learning is used where CNN layers are regularized using techniques such as adversarial training or domain adaptation losses to extract transferable features that generalize across domains (Xia, Huang, & Wang, 2020). Sequence alignment in LSTM-based temporal modeling ensures that the fault diagnosis sequence is robust to misalignments and variations in time-domain signal durations (Yang, Wan, Zhang, & Xiong, 2022).

These adaptations are rooted in the theory of domain adaptation and representation learning, which aim to minimize the discrepancy between source and target domains by learning invariant features (H. Zhao, Combes, Zhang, & Gordon, 2019).

Additionally, the proposed CNN-LSTM architecture distinguishes itself by integrating preprocessing techniques such as RMS and standardization, which are tailored to improve model stability and adaptability across diverse domains. Unlike traditional CNN-LSTM combinations, this approach leverages multimodal data fusion of vibration and acoustic signals to simultaneously extract spatial and temporal features, thereby addressing the cross-domain variability often encountered in industrial datasets.

To enhance this hybrid-model's accuracy, it is important to highlight preprocessing techniques and multimodal data fusion separately. The preprocessing phase incorporates

normalization, RMS calculations to condense vibration signals into representative metrics, effectively reducing noise and capturing the signal's overall energy. Preprocessing is used to normalize the data across all input channels, ensuring that features have a consistent scale and improving the convergence behavior during training. These steps are essential in mitigating domain-specific disparities, making the model more robust to variations in data distributions across different industrial setups.

The multimodal fusion aspect of the architecture integrates standardized vibration and acoustic signals, enabling the CNN component to focus on spatial patterns and feature hierarchies within each modality. The LSTM layers subsequently capture temporal dependencies, which are crucial for understanding sequential dynamics inherent in machine operations. By combining these, the architecture addresses the cross-domain variability effectively, as it learns to generalize feature representations that are invariant to domain-specific characteristics.

The model explicitly tackles cross-domain challenges through the novel combination of preprocessing methods and architecture design. The domain splitting methodology ensures complete separation of training and testing datasets, enhancing the evaluation's robustness. By incorporating preprocessing steps that normalize data and amplify relevant features, the model improves its ability to generalize across varying operational conditions.

The network architecture is tuned for a parameter optimization process that considers domain-specific variations, including input signal lengths and batch sizes optimized for cross-domain scenarios. This ensures that the model performs consistently, even when tested on datasets with naturally occurring faults, a key differentiator from existing approaches.

While validated on the UORED-VAFCLS dataset, the methodology uses a unique domain-splitting framework that simulates real-world cross-domain conditions. This setup, combined with the fusion of multiple data modalities, provides a practical evaluation of the model's performance, addressing the critique regarding reliance on public datasets.

3. METHODOLOGY

3.1. Statistical Methods

Bearing data preprocessing is an important task in ML for fault classification (Gupta, Wadhvani, & Rasool, 2023; B. Li, Song, & Zhao, 2024). This is especially true when data available for training algorithms has not been collected under ideal conditions. Additionally, preprocessing is expected to reduce the computational time required for detecting bearing faults. Despite the availability of various techniques, there is still a research gap in determining which statistical methods can most effectively improve the quality of ML models for sensor data, thereby ensuring the reliability and accuracy of classification models. Existing methods do not use simple

statistical measures such as standardization (Muhammad Ali & Faraj, 2014) and root mean square (RMS) (Lian et al., 2024) for normalizing datasets. This paper explores the use of basic statistical techniques to preprocess bearing data so that data fed into a ML algorithm is normalized.

The first preprocessing technique used in this paper is min-max normalization, as shown in equation (4) (Cabello-Solorzano, Ortigosa de Araujo, Peña, Correia, & J. Tallón-Ballesteros, 2023). This technique preserves the original distribution of the data while ensuring all features lie within a comparable range, which is particularly useful for models that are sensitive to input scales. However, it is important to note that min-max normalization is sensitive to outliers, as extreme values can disproportionately affect the scaling. Min-max normalization scales the data to a fixed range, between -1 and 1, as seen in equation (4).

$$x_{norm}(x) = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

where x are the individual data points that are being standardized, x_{min} and x_{max} are the minimum and maximum values within the data.

The second preprocessing technique used in this paper is normalization using standardization (Z-score normalization), as shown in equation (5) (Han et al., 2022). Standardization improves centrality in the dataset, where the data is preprocessed to have a mean of 0 and a standard deviation of 1 (Muhammad Ali & Faraj, 2014). This provides a baseline for the data, without which it is difficult to judge whether there is deviation in the data that could represent healthy or faulty states. Standardization ensures that different bearing conditions within a dataset can be compared more effectively. Standardization analysis uses sensor data and looks for a common value at its center, hence reducing the effect of outliers and noise. This normalization technique is especially important for ML models, as it ensures that features used for training will lie on the same scale, boosting the model's performance.

$$x_{std}(x) = \frac{x - \mu}{\sigma} \quad (5)$$

where μ is the mean of the dataset, and σ is the dataset's standard deviation.

The third preprocessing method used in this paper is RMS, as shown in equation (6) (Samanta & Al-balushi, 2003). RMS is a measure that considers both the magnitude and variability of data. It is useful for detecting changes and variations in vibration signals, as well as bearing failure signatures. RMS is calculated by the square root of the mean of the points squared. A method like this enhances bigger deviations, making it easier to pick out significant changes in the signal that may indicate more substantial issues. The use of RMS in preprocessing underlines the general magnitude of a signal features that are important for fault classification. RMS reduces the noise in a time-domain dataset, thereby offering more robustness against overfitting by ML models.

$$x_{RMS}(x) = \sqrt{\frac{\sum x^2}{n}} \quad (6)$$

where n is the total number of data points in the dataset.

There are various reasons why preprocessing raw bearing data using statistical methods like standardization and RMS is important. These methods make the effects of noise and outliers less impactful. Such normalization is critical in ensuring that ML models are improved in their ability to learn from the data. Preprocessing also helps to identify and eliminate inconsistencies embedded within sensor readings, often greatly impacting the performance of a fault classification model. Finally, use of simple statistical methods for preprocessing makes the process more approachable and easier to implement, with a robust base that can support the use of more complex techniques. In general, an effective preprocessing framework for bearing data is crucial for ML-based fault classification for predictive maintenance and consequently reduces downtime in industry.

3.2. ML Algorithms

CNNs are a branch of DL models designed for processing image-based data [11]. CNNs consist of multiple layers that include convolutional layers, pooling layers, and fully connected layers. Data is filtered through convolutional layers allowing the capture of patterns like textures and edges of spatial features (LeCun, Bengio, & Hinton, 2015). Pooling layers are used to reduce the dimensionality of the data, making the network more efficient for smaller inputs. This allows CNNs to be efficient at learning features from raw data (LeCun et al., 2015).

RNNs, on the other hand, use sequential data like time series. RNNs have hidden states that store information from previous time steps that allow for learning temporal dependencies in the data (LeCun et al., 2015). This makes RNNs important when the order of data points is critical. The RNN captures temporal dependencies and tracks changes over time, making it well-suited for modeling the evolution of machine states across sequential data. Nonetheless, training RNNs is challenging due to vanishing or exploding gradients that can affect learning over long sequence datasets. Therefore, more advanced variants such as LSTM networks have been developed to address these issues, allowing RNNs to capture long-term dependencies over long sequences (LeCun et al., 2015).

A combined CNN-LSTM architecture is proposed in Table 1 for classifying five classes of machine conditions: healthy, as well as ball, inner, outer, and cage faults for bearing cross domain validation using time domain accelerometer data. The algorithm is inspired by a previous condition monitoring algorithm for bearing detection proposed by Sehri et. al. (Sehri, Ertarğın, Orhan, Yildirim, & Dumond, 2024). This algorithm consists of a series of Conv2D layers with max-pooling layers that help extract hierarchical features from the input signal. The convolutional

layers all have distinct filter sizes and depths and use ReLu activation functions. These features are then passed through an LSTM layer to capture temporal dependencies in the sequential data before being flattened and fed into two linear layers to produce a probability distribution over the five bearing classes. The output is then flattened and fed into two fully connected layers, using ReLu activation to produce a probability distribution over the five bearing states. This model tries to pick out patterns from input data using convolutional operations and gets trained on categorical cross-entropy as the loss function validates different bearings of the same size with different machine states across domains.

Table 1. CNN-LSTM ARCHITECTURE

Layers	Structures
1	Conv2d(18, 32, (1, 7)), BN, ReLU
2	Conv2d(32, 64, (1, 5)), BN, ReLU, MaxPool(1, 2)
3	Conv2d(64, 128, (1, 3)), BN, ReLU
4	Conv2d(128, 256, (1, 3)), BN, ReLU, AdaptiveMaxPool(1)
5	LSTM(256, 256, 3)
6	Linear(256, 256)
7	Linear(256, N_c)

The selection of convolutional kernel sizes, number of layers, and feature depths in the CNN-LSTM architecture was determined through iterative empirical testing via an ablation study, balancing accuracy and computational efficiency for cross-domain bearing fault diagnosis. Larger kernel sizes were avoided to reduce the risk of overfitting on smaller receptive fields, while varying kernel dimensions across layers was found to enhance feature diversity for fault patterns. The depth of the architecture, including the single LSTM layer for accelerometer and acoustic data, was chosen to capture sufficient temporal dependencies without incurring excessive training time or memory requirements. This configuration offered the best trade-off between performance and resource usage in preliminary experiments, making it well-suited for the industrial fault detection scenarios considered in this study.

The same architecture is utilized for time domain acoustic data to classify healthy, ball, inner, outer, and cage faults in bearing cross domains based on a combination of CNN-LSTM layers. The architecture uses Conv2D layers with varying filter sizes and depths, followed by max-pooling layers to extract features of the data. Then, 3 LSTM layers of 256 kernels are included, which lead to a flattened layer. This is followed by two dense layers both with 256 kernels, giving the probability distribution over five classes. This model merges the advantages of both CNNs and RNNs in effectively capturing features from the accelerometer and acoustic data individually.

Vibration and acoustic data are then fused together in parallel. Because of the robustness of the CNN and LSTM layers in this multimodal deep model, as shown in Figure 1, healthy, ball, inner, outer, and cage classes are extracted using cross domain learning. Here, CNN-LSTM layers are developed for acoustic data so that computation time is reduced while maintaining high performance. This neural network architecture allows one to train both acoustic and

accelerometer data inputs at the same time with the aggregated algorithms from each individual model. By merging input signals in parallel, data should strengthen categorization results using the distinct features captured by each model for higher testing accuracy.

The framework proposed by Zhao et al. is used to create a ML model that improves time domain data to compare cross domain accuracies of bearing conditions (Z. Zhao et al., 2021). Dataset splitting in this paper uses training, validation, and test sets that do not overlap to avoid data leakage. Specifically, training and validation sets are drawn entirely from individual domains (source and target), while the test set comes from an entirely separate domain that is not used during training or validation, ensuring a robust evaluation.

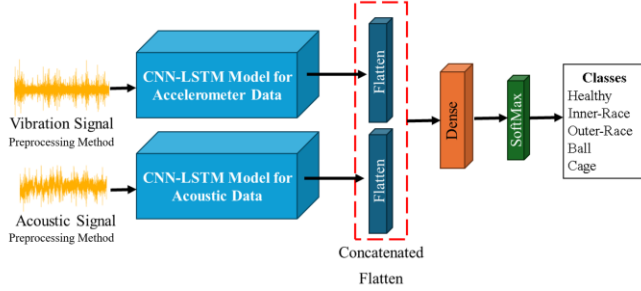


Figure 1. Proposed CNN-LSTM Model.

3.3. Datasets

For this study, the CWRU (“Download a Data File | Case School of Engineering | Case Western Reserve University,” 2021) and UORED-VAFCLS datasets are used (Sehri & Dumond, 2023; Sehri, Dumond, & Bouchard, 2023b). The UORED-VAFCLS dataset is particularly valuable due to its naturally occurring faults that develop over time, making it ideal for simulating real-world deterioration conditions and enhancing the robustness of the models. Statistical preprocessing methods, such as min-max normalization, standardization, and RMS, are applied to the sensor data to improve the detection of bearing health states. For these techniques to be effective, datasets must include diverse load conditions, speeds, or fault types. Domain splitting within the dataset further enhances the accuracy and reliability of fault classification by focusing on naturally developed fault conditions. The use of the UORED-VAFCLS dataset provides practical insights for industrial predictive maintenance. However, the dataset is limited by the fact that the sensor data is clean compared to noisy industrial environments. This study represents an initial step toward integrating ML research into industrial applications for bearing condition monitoring, with the goal of improving fault detection and predictive maintenance practices.

Table 2 and 3 show domain splitting of the CWRU and UORED-VAFCLS datasets, respectively. For the CWRU dataset, each domain consists of an inner race, outer race, and ball fault files. The UORED-VAFCLS dataset contains similar files, with the addition of healthy and cage fault files. Data sample naming conventions consist of a letter

representing the type of fault, a number representing different 6203 (UORED-VAFCLS) and 6205 (CWRU drive end) ball bearings being tested and a second number representing the bearing state (i.e., for UORED-VAFCLS: 0- healthy, 1- fault developing, and 2- faulty and for CWRU: 0, 1, 2, 3 represent different speeds at 1797, 1722, 1750, and 1730 RPM respectively). An example domain consists of files such as H-1-0, I-1-1, O-6-1, B-11-1, C-16-1 in a single domain called domain 1 (Domain Name). The proposed cross domain testing framework is conducted by combining 4 domains together and testing on another domain to see if the trained data transfers across domains for different bearings of the same size. CWRU dataset splitting is provided in Figure 2, and the dataset is organized based on load conditions represented by motor horsepower (HP): 1 HP, 2 HP, and 3 HP. Each load condition corresponds to different operating domains (Domains 1–12) and includes vibration data captured from bearings (Drive End) with varying levels of fault sizes: 0.007", 0.014", and 0.021". The figure highlights two approaches for dataset splitting: the load splitting approach, which mixes data across motor speeds, leads to data leakage while training and testing on the same bearing, and the proposed splitting approach, where the test domain mixes data across fault severity conditions and unique bearing entities, which makes this splitting strategy more akin to industrial situations where unseen bearing fault sizes and unique bearing entities are introduced (Hendriks et al., 2022). Since the data is split by fault sizes rather than loads, accuracies for this difficult task are expected to be very low according to Hendrik’s et al. (Hendriks et al., 2022).

In this work, each dataset is analyzed in a cross-domain manner within itself, meaning that some domains are allocated exclusively for training, others for validation, and the remaining for testing. This ensures that the model is evaluated on entirely unseen domains, simulating industrial generalization scenarios where fault conditions differ from those in the training set. This approach should not be confused with transfer learning, where training occurs on one dataset (e.g., CWRU) and testing is performed on a different dataset (e.g., UORED-VAFCLS) (Sehri, Varejão, et al., 2025). The use of the term “cross-domain” strictly refers to evaluating model generalization across distinct operating domains within the same dataset, not across fundamentally different datasets.

UORED-VAFCLS dataset splitting is provided in Figure 3 and Figure 4. This dataset consists of 20 different 6203 healthy bearings that are submitted to accelerated life testing so that faults are developed naturally, providing 5 separate bearings of each fault condition. Figure 3 presents vibration data from 6203 ball bearings across healthy, developing, and fully developed fault conditions, categorized into five bearing states: healthy, inner race fault, outer race fault, ball fault, and cage fault. The dataset is split into distinct domains, with no overlap between training, validation, and testing sets, simulating real-world industrial scenarios where models encounter entirely new conditions during testing. This

domain-based splitting ensures robust evaluation of generalization capabilities, reflecting industrial requirements for diagnosing faults under varying operating conditions. Features were extracted from the time domain, and statistical analyses were conducted to extract information for model training and evaluation. The splitting approach used herein allows for comparing ML models while emphasizing their applicability in scenarios where fault conditions differ significantly from the training data, thereby highlighting the industrial relevance of the proposed method.

Table 2. DOMAIN SPLITTING FOR THE CWRU DATASET 48

DRIVE END (“Download a Data File | Case School of Engineering | Case Western Reserve University,” 2021)

Domain Name	Inner Race	Outer-Race	Ball
1	I-5-0 [IR007 0]	O-8-0 [OR007@6 0]	B-11-0 [B007 0]
2	I-5-1 [IR007 1]	O-8-1 [OR007@6 1]	B-11-1 [B007 1]
3	I-5-2 [IR007 2]	O-8-2 [OR007@6 2]	B-11-2 [B007 2]
4	I-5-3 [IR007 3]	O-8-3 [OR014@6 3]	B-11-3 [B007 3]
5	I-6-0 [IR014 0]	O-9-0 [OR014@6 0]	B-12-0 [B014 0]
6	I-6-1 [IR014 1]	O-9-1 [OR014@6 1]	B-12-1 [B014 1]
7	I-6-2 [IR014 2]	O-9-2 [OR014@6 2]	B-12-2 [B014 2]
8	I-6-3 [IR014 3]	O-9-3 [OR014@6 3]	B-12-3 [B014 3]
9	I-7-0 [IR021 0]	O-10-0 [OR021@6 0]	B-13-0 [B021 0]
10	I-7-1 [IR021 1]	O-10-1 [OR021@6 1]	B-13-1 [B021 1]
11	I-7-2 [IR021 2]	O-10-2 [OR021@6 2]	B-13-2 [B021 2]
12	I-7-3 [IR021 3]	O-10-3 [OR021@6 3]	B-13-3 [B021 3]

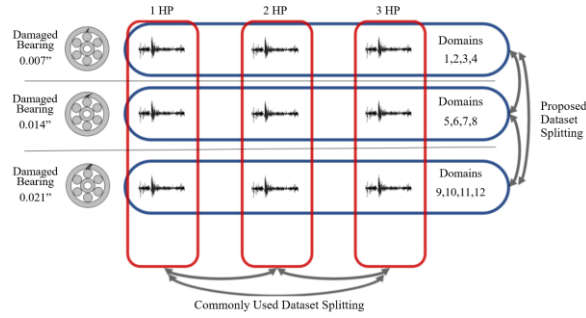


Figure 2. Proposed CWRU Dataset Splitting (Hendriks et al., 2022).

Table 3. DOMAIN SPLITTING FOR THE UORED-VAFCLS DATASET

Domain Name	Healthy	Inner-Race	Outer-Race	Ball	Cage
1	H-1-0	I-1-1	O-6-1	B-11-1	C-16-1
2	H-2-0	I-1-2	O-6-2	B-11-2	C-16-2
3	H-3-0	I-2-1	O-7-1	B-12-1	C-17-1
4	H-4-0	I-2-2	O-7-2	B-12-2	C-17-2
5	H-5-0	I-3-1	O-8-1	B-13-1	C-18-1
6	H-6-0	I-3-2	O-8-2	B-13-2	C-18-2
7	H-7-0	I-4-1	O-9-1	B-14-1	C-19-1
8	H-8-0	I-4-2	O-9-2	B-14-2	C-19-2
9	H-9-0	I-5-1	O-10-1	B-15-1	C-20-1
10	H-10-0	I-5-2	O-10-2	B-15-2	C-20-2

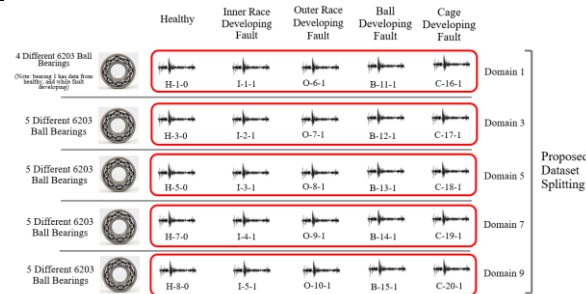


Figure 3. Proposed UORED-VAFCLS Dataset Splitting for Healthy and Developing Fault Bearings

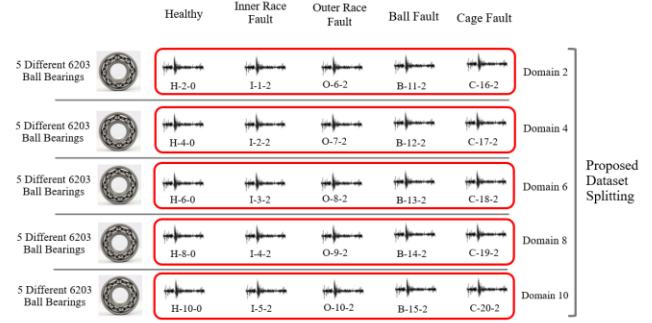


Figure 4. Proposed UORED-VAFCLS Dataset Splitting for Healthy and Faulty Bearings

4. EXPERIMENTAL VERIFICATION

The results consist of testing different domains using cross domain learning, with both the CWRU and UORED-VAFCLS datasets. Each test is run 10 times and the average accuracy of the best epoch for each test is taken. A total of 100 epochs are evaluated, and the learning rate is set to 0.001 for each case. Additionally, the experiments were conducted using Python 3.10.9 and PyTorch 1.12.0 installed via conda. All ML architectures were trained on an NVIDIA GeForce RTX 3070 GPU with CUDA version 11.3 and CUDNN 8.1. The machine used for experiments ran on Windows 11 with an Intel Core i9-12900H CPU, 1TB SSD, and 32 GB of RAM. Key Python libraries used include numpy 1.23.4, matplotlib 3.6.3, pandas 1.5.0, scikit-learn 1.1.2, torch 2.3.1+cu118, tqdm 4.64.1, torchvision 0.18.1+cu118, torchaudio 2.3.1+cu118, datasets 2.19.1.

In this setup, the PC served as the primary workstation for orchestrating and cycling through different model designs, with each design being trained on the NVIDIA GPU.

4.1. Accelerometer Data Results

Table 4 demonstrates the performance of different ML architectures and hyperparameters when used with the CWRU dataset’s accelerometer data. Domains 1 to 8 are used for training and validation while domain 9 is used for testing. The proposed 1D CNN-LSTM model exhibits fair stability with test accuracies exceeding 60% in majority of the cases and a low standard deviation, particularly when using RMS preprocessing. Optimal performance was achieved with a batch size of 128 and an input signal length of 1024, reaching a test accuracy of 64.04 ± 2.45 . These findings suggest that the model configuration with RMS preprocessing, a batch size of 128, and an input signal length of 1024 should be utilized for further accelerometer data testing, as it offers the best combination of accuracy and stability. These results demonstrate an improvement over traditional methods when used with the more difficult data splitting framework proposed by Hendriks et al. (Hendriks et al., 2022), justifying the use of the combined CNN-LSTM architecture. This highlights the potential benefits of integrating CNNs with

RNNs with statistical preprocessing for improved fault detection performance with the CWRU dataset. Other benefits include reduced computational times while maintaining high fault detection accuracies for potential future implementation in industry.

Table 4. ACCELEROMETER RESULTS FOR DIFFERENT ML MODELS AND HYPERPARAMETERS FOR THE CWRU DATASET(“Download a Data File | Case School of Engineering | Case Western Reserve University,” 2021)

Model Type	Preprocessing	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Batch Size	Input Signal
1D CNN	Min-Max Scaling	99.65	98.54	53.11±2.52	64	512
		99.08	99.62	52.88±3.44	64	1024
		99.24	98.57	55.41±3.96	64	2048
		99.82	99.03	52.91±3.07	128	512
		99.77	99.13	53.57±2.34	128	1024
		99.33	99.37	56.78±2.91	128	2048
	Standardization	99.23	99.42	54.34±3.52	64	512
		99.14	99.01	54.88±2.30	64	1024
		99.65	99.82	59.04±3.96	64	2048
		99.33	99.31	55.70±3.93	128	512
		99.53	99.55	54.66±3.40	128	1024
		99.86	99.94	57.40±3.45	128	2048
	RMS	99.55	99.81	53.85±1.67	64	512
		99.31	99.01	54.31±3.34	64	1024
		99.92	99.98	61.43±3.88	64	2048
		99.56	97.09	55.86±2.91	128	512
		99.97	99.59	58.04±2.36	128	1024
		99.65	99.48	57.84±3.92	128	2048
1D CNN-LSTM	Min-Max Scaling	99.05	99.83	54.23±1.96	64	512
		99.82	99.01	55.27±2.34	64	1024
		99.37	99.61	56.94±3.78	64	2048
		99.61	99.33	54.33±2.67	128	512
		99.32	99.64	57.19±2.89	128	1024
		99.89	99.06	55.72±2.11	128	2048
	Standardization	99.88	99.94	56.10±1.70	64	512
		99.54	99.77	57.34±1.73	64	1024
		99.75	99.88	58.12±4.58	64	2048
		98.75	99.46	56.05±1.59	128	512
		99.66	99.72	58.06±3.07	128	1024
		99.89	99.35	58.73±3.81	128	2048
	RMS	99.33	99.35	55.77±1.33	64	512
		99.51	99.47	62.63±2.33	64	1024
		99.62	99.77	63.66±4.24	64	2048
		99.53	99.09	61.00±2.78	128	512
		99.58	99.32	64.04±2.45	128	1024
		99.66	99.47	63.25±3.67	128	2048

Table 5 shows the results when using the UORED-VAFCLS dataset for validation domains 2, 4, 6, and 8 being tested on domain 10 using cross domain analysis of the bearing accelerometer dataset to identify the best hyperparameters. Odd numbered domains include developing fault data, whereas even numbered domains include fully developed fault data. The reason even-numbered domains alone are selected for testing hyperparameters is that developing fault and faulty bearing data is found to be too similar, which biases the results. This approach serves as a first step in addressing the detection challenge. Once high accuracy is achieved with these even-numbered domains, the next step will be to combine all domains for more comprehensive testing and improved fault detection performance. Table 5 demonstrates that the proposed 1D CNN-LSTM model has a test accuracy above 90% and a standard deviation of less than 1 when using RMS preprocessing, indicating that the proposed model is stable. Additionally, the best accuracy is obtained using a batch size of 128 and an input signal length of 1024. Therefore, the remainder of accelerometer data testing will be done using

this model (i.e., RMS preprocessing, batch size of 128, and input signal of 1024).

Table 5. ACCELEROMETER RESULTS FOR DIFFERENT ML MODELS AND HYPERPARAMETERS FOR THE UORED-VAFCLS DATASETS

Model Type	Preprocessing	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Batch Size	Input Signal
1D CNN	Min-Max Scaling	99.20	99.67	76.69±2.67	64	512
		99.58	99.37	75.97±1.91	64	1024
		99.32	99.61	76.83±1.82	64	2048
		99.68	99.31	78.41±3.56	128	512
		99.02	99.39	77.35±3.08	128	1024
		99.91	98.37	74.70±1.55	128	2048
	Standardization	99.76	99.94	79.33±1.91	64	512
		99.54	99.88	79.50±1.80	64	1024
		99.98	99.82	79.27±2.16	64	2048
		99.92	99.82	80.95±1.96	128	512
		99.98	99.13	80.56±2.71	128	1024
		99.74	99.94	79.27±2.01	128	2048
	RMS	99.42	99.94	78.42±3.18	64	512
		99.83	99.76	78.70±2.86	64	1024
		99.95	99.94	79.55±3.96	64	2048
		99.98	99.94	79.69±3.47	128	512
		99.65	99.94	78.93±3.02	128	1024
		99.91	99.82	79.79±3.58	128	2048
1D CNN-LSTM	Min-Max Scaling	99.86	99.07	85.18±2.32	64	512
		98.34	99.61	87.55±1.09	64	1024
		99.30	99.05	83.15±1.67	64	2048
		99.73	99.88	85.97±2.15	128	512
		99.11	98.84	87.17±1.63	128	1024
		99.17	99.35	84.31±2.97	128	2048
	Standardization	99.89	99.76	87.49±1.23	64	512
		99.76	99.82	88.61±1.29	64	1024
		99.22	99.88	86.44±0.94	64	2048
		99.38	99.21	89.46±1.64	128	512
		99.81	99.82	90.50±0.88	128	1024
		99.08	99.88	89.76±1.04	128	2048
	RMS	99.82	99.97	92.93±0.94	64	512
		99.12	99.94	92.83±0.89	64	1024
		99.92	99.94	92.74±0.97	64	2048
		99.98	99.94	92.19±0.86	128	512
		99.86	99.15	93.20±0.73	128	1024
		99.47	99.94	92.03±0.87	128	2048

The results in Table 6 highlight the challenges of cross-domain fault diagnosis, particularly when handling data with different fault severities. The table compares existing results of traditional cross domain methods, such as the atrous convolutional deep inception network (ACDIN), the wide first-layer kernel deep convolutional neural networks (WDCNN), AlexNet, k-nearest neighbour (KNN), and support vector machine (SVM), all of which employ different preprocessing techniques like the fast Fourier transform (FFT) and spectrograms. Additionally, advanced cross domain methods, such as multi-kernel maximum mean discrepancy (MK-MMD) (J. Li et al., 2024), the domain adversarial network (DANN) (Zhuorui Li, Ma, Wu, Li, & Wang, 2024), and transformer networks (Mirzaebonehkhater, Labbaf-Khaniki, & Manthouri, 2024) are tested for comparison. Despite the variety of approaches, these models consistently achieve test accuracies below 60% for the more difficult task of domain splitting across fault severities, as shown in the literature (Hendriks et al., 2022; Rauber et al., 2021). This more difficult data splitting task, based on different fault sizes across domains, is closer to what would be observed in industry. In contrast, the CNN-LSTM model achieves an average test accuracy of 61.38 ± 2.42 ,

Table 6. ACCELEROMETER RESULTS FOR EXISTING AND SELECTED HYPERPARAMETERS ON DIFFERENT FAULT SEVERITY DOMAINS OF THE CWRU

Cross Domain Methods	Model Type	Preprocessing	Domain Train Validation	Domain Tested	Test Accuracy (%)	Batch Size	Input Signal	Computational Time (seconds)
Traditional	ACDIN*	Raw	D, E, F	D, E, F	37.13 ± 12.70	N/A	5118	N/A
	ACDIN*	FFT			40.57 ± 11.01	N/A	10,236	N/A
	WDCNN*	Raw			36.20 ± 9.49	N/A	2048	N/A
	WDCNN*	FFT			50.02 ± 6.39	N/A	4096	N/A
	AlexNet*	Spectrogram			47.37 ± 11.03	N/A	11,500	N/A
	ResNet*	Spectrogram	Average of 0.007, 0.014, 0.021	0.007, 0.014, 0.021	53.10 ± 6.57	N/A	11,500	N/A
	KNN**	Raw			50.72 ± 1.28	N/A	N/A	N/A
	SVM**	Raw			49.76 ± 2.54	N/A	N/A	N/A
	RF**	Raw			52.56 ± 1.13	N/A	N/A	N/A
	1D CNN**	Raw			53.26 ± 1.90	N/A	N/A	N/A
Advanced	CNN-LSTM	RMS	Average of 0.007, 0.014, 0.021	0.007, 0.014, 0.021	61.38 ± 2.42	128	1024	55
	MK-MMD	RAW			57.34 ± 2.89	128	1024	80
	DANN	RAW			56.65 ± 3.73	128	1024	75
	Transformer	RAW			57.02 ± 3.31	128	1024	90

demonstrating its relative improvement over both traditional and recent advanced methods.

Unlike existing approaches, the proposed model integrates preprocessing using RMS, which contributes to stabilizing performance across domains. This preprocessing step enhances the model's ability to generalize features, reducing the impact of domain-specific noise and inconsistencies. For example, while network models in prior studies achieve average accuracies below 55% on the same dataset across fault severities, the proposed model exceeds this, achieving accuracies above 60% in most domains. This improvement underscores the benefits of integrating spatial and temporal feature extraction with preprocessing.

* (Hendriks et al., 2022), ** (Raubert et al., 2021)

Figure 5 visually represents the comparative performance of these models, as summarized in Table 6. The bar chart demonstrates that traditional approaches such as ACDIN, WDCNN, and AlexNet using spectrograms as an input alone exhibit a wide range of average test accuracies, with standard deviation bars indicating the variability in their performance while advanced methods perform slightly better than most traditional methods in terms of test accuracy. Additionally, when computational times in Table 6 for the CNN-LSTM method are compared to the advanced methods, the CNN-LSTM combined with statistical preprocessing outperforms all three. The figure highlights that while certain

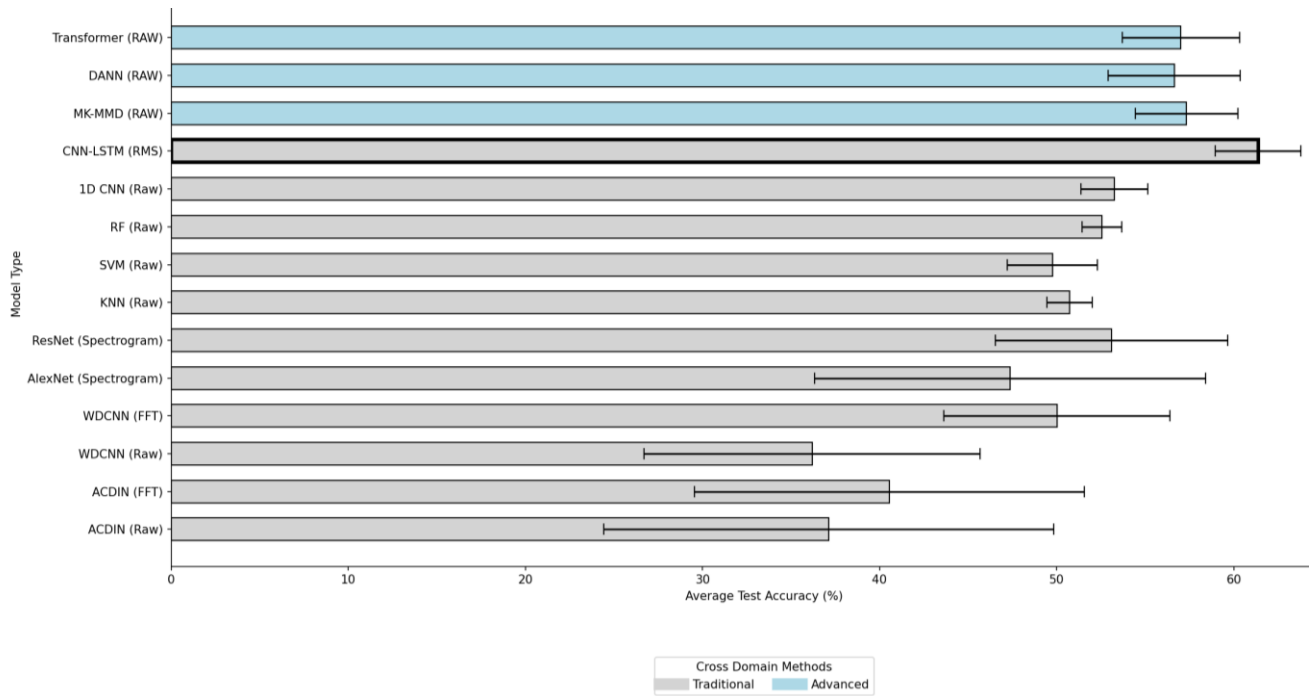


Figure 5. Average Test Accuracy When Splitting Data by Fault Severities with the CWRU Dataset

methods like ResNet with spectrogram inputs reach a test accuracy of 53.10 ± 6.57 , none surpass the 55% accuracy consistently across domains. On the other hand, the proposed CNN-LSTM model, in the figure, clearly stands out with a significantly higher accuracy of 61.38 ± 2.42 , validating the effectiveness of its RMS preprocessing approach and the integration of convolutional and sequential layers for spatial and temporal feature extraction. The figure shows the improvement in test accuracy by the proposed model and aligns with the numerical findings, visually reinforcing the advantages of the CNN-LSTM methodology.

The domain-splitting methodology and multimodal data fusion further contribute to experimental data innovation by addressing challenges that other methods in Table 6 fail to overcome. The strict separation of training and testing domains ensures that the evaluation simulates realistic cross-domain conditions, validating the model's effectiveness. This approach sets a foundation for future research to refine feature extraction techniques and address residual challenges, as seen in domains with lower performance. The results indicate that the proposed method effectively leverages preprocessing steps alongside algorithmic integration to handle fault severities, which has not been explored extensively in prior studies.

The results in Table 6 and the average accuracies provided in the existing literature further emphasize the strengths and limitations of the proposed approach. The CNN-LSTM model achieves improvements, with average test accuracies exceeding 60%, showing that traditional methods when combined with statistical preprocessing can surpass advanced cross domain methods. This disparity underlines how traditional cross domain methods can enhance generalization and consistency across different testing conditions and result in lower computational time.

4.2. Acoustic Data Results

Table 7 shows the results for the UORED-VAFCLS acoustic bearing dataset. Domains 2, 4, 6, and 8 are used for training and validation while domain 10 is used for testing. The results are similar to the accelerometer data, so for the remainder of the testing on acoustic data, a 1D CNN-LSTM model with RMS preprocessing, a batch size of 128, and an input signal length of 1024 are used.

Table 7. ACOUSTIC RESULTS FOR DIFFERENT ML MODELS AND HYPERPARAMETERS FOR THE UORED-VAFCLS

DATASETS						
Model Type	Preprocessing	Train Accuracy (%)	Validation Accuracy (%)	Test Accuracy (%)	Batch Size	Input Signal
1D CNN	Min-Max Scaling	99.17	99.33	77.62±1.84	64	512
		98.29	99.08	76.94±3.21	64	1024
		99.57	99.63	78.31±2.34	64	2048
		99.93	99.37	79.13±2.74	128	512
		99.74	99.76	77.67±1.87	128	1024
	Standardization	99.07	99.91	76.09±1.99	128	2048
		99.66	99.08	80.80±2.93	64	512
		99.79	99.83	79.83±2.73	64	1024
		99.90	99.42	79.49±1.91	64	2048
		99.12	99.02	80.50±1.70	128	512
	RMS	99.67	98.90	79.10±0.96	128	1024
		99.21	99.88	79.27±2.18	128	2048
		99.20	99.53	79.70±2.50	64	512
		98.32	99.11	78.43±1.00	64	1024
		99.31	98.65	80.88±3.22	64	2048
		99.01	99.37	80.01±1.62	128	512
		99.84	99.71	78.49±2.02	128	1024
		99.82	99.23	79.98±3.92	128	2048
		99.71	99.18	86.35±2.06	64	512
		99.36	99.74	87.64±1.87	64	1024
1D CNN-LSTM	Min-Max Scaling	99.87	99.46	88.37±1.97	64	2048
		99.37	99.10	85.19±2.33	128	512
		99.03	99.74	87.67±2.09	128	1024
		99.97	99.06	85.94±1.98	128	2048
		99.10	99.32	89.29±1.79	64	512
	Standardization	99.73	99.19	90.24±1.57	64	1024
		99.23	99.51	90.36±1.49	64	2048
		99.82	99.14	88.38±1.99	128	512
		99.56	99.70	90.35±2.41	128	1024
		99.66	99.92	87.95±1.65	128	2048
	RMS	99.96	99.31	94.11±1.72	64	512
		99.57	99.81	94.52±1.20	64	1024
		99.20	99.11	94.53±0.78	64	2048
		99.84	99.45	94.29±0.70	128	512
		99.46	99.73	94.58±0.65	128	1024
		99.70	99.88	94.39±0.79	128	2048

4.3. Ablation Study for the UORED-VAFCLS Dataset

Ablation experiments are performed to verify the effectiveness of the proposed method. Table 8 summarizes the experiments conducted, where model components and configurations are analyzed for the UORED-VAFCLS dataset. By removing specific network layers, preprocessing steps, and selecting a sensor, certain features are excluded. Figure 6 provides domain test accuracy results for seven methods (A1–A8) across 10 test domains. It can be observed that the proposed method (M9) achieves the highest average accuracy of $79.13 \pm 0.92\%$, outperforming other methods. For instance, A2 and A6 show the lowest accuracies (63.13 ± 2.00 and 61.76 ± 1.67 , respectively), indicating that removing key components negatively affects the model's performance.

Figure 6 presents a box plot summarizing the accuracy

Table 8. ABLATION EXPERIMENTS FOR MODEL COMPONENTS AND CONFIGURATION OF THE UORED-VAFCLS DATASET

Method	CNN Layers	LSTM Layers	RMS Preprocessing	Accelerometer Data	Acoustic Data	Note
A1	✓		✓	✓		w single data
A2	✓		✓		✓	w single data
A3		✓	✓	✓		w single data
A4		✓	✓		✓	w single data
A5	✓	✓			✓	w/o preprocess
A6	✓	✓		✓		w/o preprocess
A7	✓	✓	✓	✓		w single data
A8	✓	✓	✓		✓	w single data
M9	✓	✓	✓	✓	✓	Proposed

distributions of each method with standard deviations. The proposed method (M9) stands out, achieving both high median accuracy and a narrower variance compared to other methods. This highlights the robustness and effectiveness of the proposed configuration. It should be noted that M9 corresponds to the fusion model, as shown in Table 8, where both the accelerometer and acoustic data columns are checked to indicate the inclusion of fusion. Conversely, methods such as A2, A3, and A6 exhibit broader variance and lower median values, emphasizing their lack of consistency and reduced generalization. The results clearly demonstrate that excluding certain components significantly degrades performance, while the inclusion of all features in the proposed method enhances generalization and accuracy.

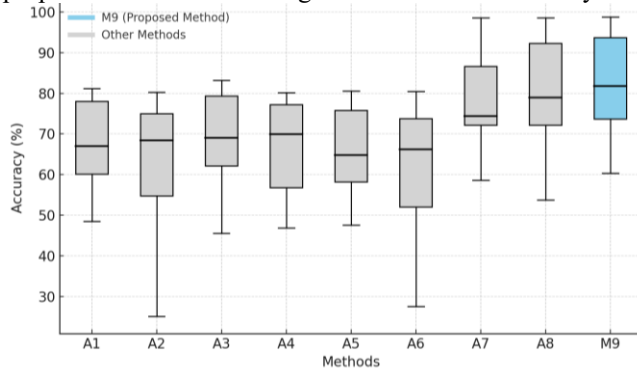


Figure 6. Box Plot of Ablation Study Results with Standard Deviations by Method with the UORED-VAFCLS Dataset

5. CONCLUSION

This study demonstrates that, if applied correctly, traditional cross domain strategies such as CNN-LSTM networks combined with simple statistical preprocessing techniques can outperform advanced cross domain algorithms for fault severity-based bearing fault diagnosis on the CWRU dataset. Additionally, by combining preprocessed time-domain data from both acoustic and accelerometer signals, the proposed methodology demonstrates improvements in fault detection accuracy and model stability on the UORED-VAFCLS dataset when compared to 1D-CNNs alone, especially for more difficult cross-domain analysis tasks. The fusion of multi-source data further enhances the reliability and robustness of the fault diagnosis process. Further refinement of this methodology could provide improvements for these tasks.

The proposed method has been compared with the available literature for the CWRU dataset. Nonetheless, future work will focus on optimizing the model to achieve even higher accuracy and stability while minimizing the amount of data required for effective condition monitoring. Future work will extend the comparative analysis to include advanced architectures such as Transformers with temporal embedding and temporal convolutional networks, as well as more sophisticated fusion techniques to better exploit interdependencies between acoustic and vibration data.

Interpretability frameworks like SHAP and LIME will also be integrated to enhance model transparency and industrial applicability. As such, the development of these methods has the potential to reduce machinery downtime and improve operational efficiency in industrial settings.

NOMENCLATURE

<i>ACDIN</i>	<i>atrous convolutional deep inception network</i>
<i>AI</i>	<i>artificial intelligence</i>
<i>CNN</i>	<i>convolutional neural network</i>
<i>CWRU</i>	<i>case western reserve university</i>
<i>DANN</i>	<i>domain adversarial network</i>
<i>DL</i>	<i>deep learning</i>
<i>KNN</i>	<i>k-nearest neighbor</i>
<i>LSTM</i>	<i>long short term memory</i>
<i>ML</i>	<i>machine learning</i>
<i>MK-MMD</i>	<i>multi-kernel maximum mean discrepancy</i>
<i>PHM</i>	<i>prognostic health management</i>
<i>RNN</i>	<i>recurrent neural network</i>
<i>SVM</i>	<i>support vector machine</i>
<i>UORED-VAFCLS</i>	<i>university of ottawa rolling-element dataset- vibration and acoustic faults under constant load and speed conditions</i>
<i>WDCNN</i>	<i>wide first-layer kernel deep convolutional neural networks</i>

REFERENCES

- Bayoudh, K. (2024). A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges. *Information Fusion*, 105, 102217. <https://doi.org/10.1016/j.inffus.2023.102217>
- Cabello-Solorzano, K., Ortigosa de Araujo, I., Peña, M., Correia, L., & J. Tallón-Ballesteros, A. (2023). The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis. In P. García Bringas, H. Pérez García, F. J. Martínez de Pisón, F. Martínez Álvarez, A. Troncoso Lora, Á. Herrero, ... E. Corchado (Eds.), *18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023)* (pp. 344–353). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42536-3_33
- Chen, Z., He, G., Li, J., Liao, Y., Gryllias, K., & Li, W. (2020). Domain Adversarial Transfer Network for Cross-Domain Fault Diagnosis of Rotary Machinery. *IEEE Transactions on Instrumentation and Measurement*, 69(11), 8702–8712. <https://doi.org/10.1109/TIM.2020.2995441>
- Download a Data File | Case School of Engineering | Case Western Reserve University. (2021, August 10). Retrieved January 15, 2024, from Case School of Engineering website:

- <https://engineering.case.edu/bearingdatacenter/download-data-file>
- Gupta, M., Wadhvani, R., & Rasool, A. (2023). A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network. *Knowledge-Based Systems*, 259, 110070. <https://doi.org/10.1016/j.knosys.2022.110070>
- Han, B., Yang, Z., Zhang, Z., Bao, H., Wang, J., Liu, Z., & Li, S. (2022). A novel rolling bearing fault diagnosis method based on generalized nonlinear spectral sparsity. *Measurement*, 198, 111131. <https://doi.org/10.1016/j.measurement.2022.111131>
- Hendriks, J., Dumond, P., & Knox, D. A. (2022). Towards better benchmarking using the CWRU bearing fault dataset. *Mechanical Systems and Signal Processing*, 169, 108732. <https://doi.org/10.1016/j.ymssp.2021.108732>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jha, R. K., & Swami, P. D. (2021). Fault diagnosis and severity analysis of rolling bearings using vibration image texture enhancement and multiclass support vector machines. *Applied Acoustics*, 182, 108243. <https://doi.org/10.1016/j.apacoust.2021.108243>
- Layeghy, S., & Portmann, M. (2023). Explainable Cross-domain Evaluation of ML-based Network Intrusion Detection Systems. *Computers and Electrical Engineering*, 108, 108692. <https://doi.org/10.1016/j.compeleceng.2023.108692>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86, 2278–2324. <https://doi.org/10.1109/5.726791>
- Lee, J., & Su, H. (2025). Rethinking industrial artificial intelligence: A unified foundation framework. *International Journal of AI for Materials and Design*, 2(2), 56–68. <https://doi.org/10.36922/IJAMD025080006>
- Li, B., Song, P., & Zhao, C. (2024). Fusing consensus knowledge: A federated learning method for fault diagnosis via privacy-preserving reference under domain shift. *Information Fusion*, 106, 102290. <https://doi.org/10.1016/j.inffus.2024.102290>
- Li, J., Ye, Z., Gao, J., Meng, Z., Tong, K., & Yu, S. (2024). Fault transfer diagnosis of rolling bearings across different devices via multi-domain information fusion and multi-kernel maximum mean discrepancy. *Applied Soft Computing*, 159, 111620. <https://doi.org/10.1016/j.asoc.2024.111620>
- Li, Zhuorui, Ma, J., Wu, J., Li, X., & Wang, X. (2024). Implicit Discriminator Domain Adversarial Residual Network for Cross Domain Rolling Bearing Fault Diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 73, 1–11. <https://doi.org/10.1109/TIM.2024.3436093>
- Li, Zixian, Ding, X., Song, Z., Wang, L., Qin, B., & Huang, W. (2024). Digital twin-assisted dual transfer: A novel information-model adaptation method for rolling bearing fault diagnosis. *Information Fusion*, 106, 102271. <https://doi.org/10.1016/j.inffus.2024.102271>
- Lian, C., Zhao, Y., Shao, J., Sun, T., Dong, F., Ju, Z., ... Shan, P. (2024). CFI-LFENet: Infusing cross-domain fusion image and lightweight feature enhanced network for fault diagnosis. *Information Fusion*, 104, 102162. <https://doi.org/10.1016/j.inffus.2023.102162>
- Lundström, A., & O’Nils, M. (2023). Factory-Based Vibration Data for Bearing-Fault Detection. *Data*, 8(7), 115. <https://doi.org/10.3390/data8070115>
- McCarter, D. (2023). Mathematical Analysis of Convolutional Neural Networks. *Dissertations and Theses*. Retrieved from <https://red.library.usd.edu/diss-thesis/114>
- Meng, Z., Zhan, X., Li, J., & Pan, Z. (2018). An enhancement denoising autoencoder for rolling bearing fault diagnosis. *Measurement*, 130, 448–454. <https://doi.org/10.1016/j.measurement.2018.08.010>
- Mirzaeibonehkhater, M., Labbaf-Khaniki, M. A., & Manthouri, M. (2024, December 15). *Transformer-Based Bearing Fault Detection using Temporal Decomposition Attention Mechanism*. arXiv. <https://doi.org/10.48550/arXiv.2412.11245>
- Muhammad Ali, P., & Faraj, R. (2014). *Data Normalization and Standardization: A Technical Report*. <https://doi.org/10.13140/RG.2.2.28948.04489>
- Rauber, T. W., da Silva Loca, A. L., Boldt, F. de A., Rodrigues, A. L., & Varejão, F. M. (2021). An experimental methodology to evaluate machine learning methods for fault diagnosis based on vibration signals. *Expert Systems with Applications*, 167, 114022. <https://doi.org/10.1016/j.eswa.2020.114022>
- Samanta, B., & Al-balushi, K. R. (2003). ARTIFICIAL NEURAL NETWORK BASED FAULT DIAGNOSTICS OF ROLLING ELEMENT BEARINGS USING TIME-DOMAIN FEATURES. *Mechanical Systems and Signal Processing*, 17(2), 317–328. <https://doi.org/10.1006/mssp.2001.1462>
- Sehri, M., & Dumond, P. (2023). *University of Ottawa Rolling-element Dataset – Vibration and Acoustic Faults under Constant Load and Speed conditions (UORED-VAFCLS)*. 5. <https://doi.org/10.17632/y2px5tg92h.5>

- Sehri, M., Dumond, P., & Bouchard, M. (2023, November 21). *Design of a Data Fusion and Deep Transfer Learning Test Rig for Roller Bearings Diagnosis and Prognosis*. Presented at the ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers Digital Collection. <https://doi.org/10.1115/DETC2023-117264>
- Sehri, M., Dumond, P., & Bouchard, M. (2023). University of Ottawa constant load and speed rolling-element bearing vibration and acoustic fault signature datasets. *Data in Brief*, 49, 109327. <https://doi.org/10.1016/j.dib.2023.109327>
- Sehri, M., Ertarğın, M., Orhan, A., Yildirim, Ö., & Dumond, P. (2024). *Deep Learning Approach to Bearing and Induction Motor Fault Diagnosis via Data Fusion*.
- Sehri, M., Hua, Z., Boldt, F. de A., & Dumond, P. (2025). Selective embedding for deep learning. *Knowledge-Based Systems*, 330, 114535. <https://doi.org/10.1016/j.knosys.2025.114535>
- Sehri, M., Varejão, I., Hua, Z., Bonella, V., Santos, A., Boldt, F. de A., ... Varejão, F. M. (2025). Towards a Universal Vibration Analysis Dataset: A Framework for Transfer Learning in Predictive Maintenance and Structural Health Monitoring. *International Journal of Prognostics and Health Management*, 16(3). <https://doi.org/10.36001/ijphm.2025.v16i3.4239>
- Su, H., & Lee, J. (2024). Machine Learning Approaches for Diagnostics and Prognostics of Industrial Systems Using Open Source Data from PHM Data Challenges: A Review. *International Journal of Prognostics and Health Management*, 15(2). <https://doi.org/10.36001/ijphm.2024.v15i2.3993>
- Wan, H., Guo, S., Yin, K., Liang, X., & Lin, Y. (2020). CTS-LSTM: LSTM-based neural networks for correlated time series prediction. *Knowledge-Based Systems*, 191, 105239. <https://doi.org/10.1016/j.knosys.2019.105239>
- Wang, H., Liu, Z., Peng, D., Yang, M., & Qin, Y. (2022). Feature-Level Attention-Guided Multitask CNN for Fault Diagnosis and Working Conditions Identification of Rolling Bearing. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9), 4757–4769. <https://doi.org/10.1109/TNNLS.2021.3060494>
- Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access*, 8, 56855–56866. <https://doi.org/10.1109/ACCESS.2020.2982225>
- Xu, M., Yu, Q., Chen, S., & Lin, J. (2024). Rolling Bearing Fault Diagnosis Based on CNN-LSTM with FFT and SVD. *Information*, 15(7), 399. <https://doi.org/10.3390/info15070399>
- Yang, X., Wan, C., Zhang, T., & Xiong, Z. (2022). Feature Extraction of Sequence Data Based on LSTM and its Application to Fault Diagnosis of Industrial Process. *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)*, 693–698. <https://doi.org/10.1109/DDCLS55054.2022.9858505>
- Yin, K., Chen, C., Luo, B., & Deng, J. (2023). A Bearing Fault Feature Cross-Domain Transfer Method Based on Motor Current Signals. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–12. <https://doi.org/10.1109/TIM.2023.3323048>
- Yu, Y., Karimi, H. R., Shi, P., Peng, R., & Zhao, S. (2024). A new multi-source information domain adaption network based on domain attributes and features transfer for cross-domain fault diagnosis. *Mechanical Systems and Signal Processing*, 211, 111194. <https://doi.org/10.1016/j.ymssp.2024.111194>
- Zhao, H., Combes, R. T. des, Zhang, K., & Gordon, G. J. (2019, May 30). *On Learning Invariant Representation for Domain Adaptation*. arXiv. <https://doi.org/10.48550/arXiv.1901.09453>
- Zhao, Z., Zhang, Q., Yu, X., Sun, C., Wang, S., Yan, R., & Chen, X. (2021). Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–28. <https://doi.org/10.1109/TIM.2021.3116309>