machines

MDPI

Article

An Improved Hybrid Lightweight Approach for Bearing Fault Detection and Classification in Three-Phase Squirrel Cage Induction Motors

Muhammad Amir Khan [1], Bilal Asad [2,*], Muhammad Usman Naseer [2], Toomas Vaimann [2] and Ants Kallaste [2]

[1] Department of Electrical Power Engineering, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[2] Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology, 19086 Tallinn, Estonia
* Correspondence: bilal.asad@taltech.ee

Abstract

Early and reliable detection of bearing faults is essential for ensuring the safe and efficient operation of rotating electrical machines, especially under varying loads and non-stationary operating conditions. However, traditional diagnostic approaches struggle to maintain accuracy when signals are noisy, high-dimensional, or affected by multiple fault patterns. To address these issues, this work presents RNN-XBoostNet, a lightweight hybrid framework that combines the temporal-feature extraction capability of Recurrent Neural Networks (RNNs) with the robust classification strength of XGBoost. A new feature-selection strategy, CoLaR-FS (integrating correlation analysis, Lasso regularization, and recursive feature elimination), is introduced to reduce redundancy and retain only the most discriminative fault features. The proposed framework is evaluated using the widely known CWRU dataset and a newly developed induction-machine dataset containing ten fault categories, including six newly introduced real-world conditions. Experimental results show significant performance improvements: accuracy increased from 87.01% to 99.35% on the CWRU dataset and from 79.98% to 99.57% on the laboratory dataset. The combination of high accuracy, reduced complexity, and strong generalization demonstrates that RNN-XBoostNet, supported by CoLaR-FS, is a practical and effective solution for modern condition-based monitoring systems.

Keywords: bearing fault diagnosis; deep learning; machine learning; feature selection; CoLaR-FS; RNN-XBoostNet; Recurrent Neural Network; XGBoost; hybrid learning; induction machine; fault detection and classification; industrial monitoring

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 19 November 2025
Revised: 22 December 2025
Accepted: 30 December 2025
Published: 5 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Research in predictive maintenance for electrical machines has grown rapidly in recent years as these systems become increasingly embedded across industrial processes. Unpredictable abnormalities in motors or generators can cause high repair costs, production losses, and even safety and efficiency concerns, which highlights the need for improvements in early-fault detection methods [1,2]. Rotating electrical machines, being widely employed in manufacturing, transport, energy, automotive, and robotics, operate on electromagnetic principles to convert electrical and mechanical energy in either direction [3,4]. Among their key components, bearings support rotating shafts, reduce friction, and protect against

mechanical wear [5]. However, bearings deteriorate due to issues such as improper lubrication, misalignment, overload, and corrosion. Common bearing defects include inner-race, outer-race, and rolling-element faults, in addition to lubrication-related problems [6]. These faults often manifest through changes in vibration, acoustic noise, temperature, or current. Early and accurate identification of such anomalies is essential for preventing severe damage, avoiding unplanned downtime, and improving maintenance planning [7]. Figure 1 presents typical bearing fault types and their contribution to induction motor failures.

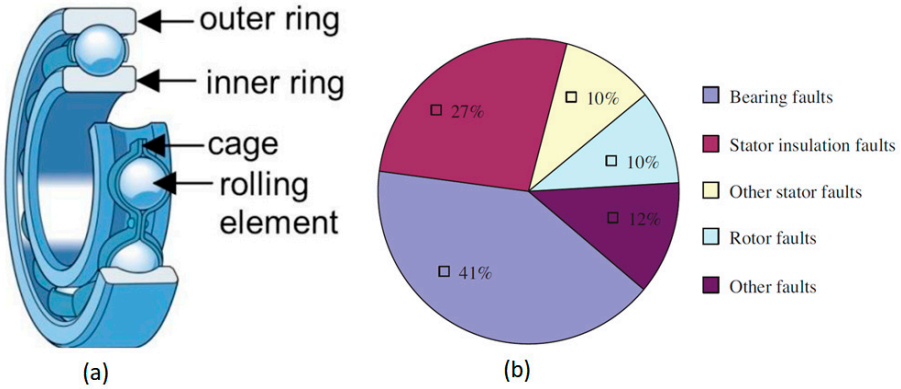


Figure 1. (**a**) Schematic diagram of bearing with faults. (**b**) Percentage distribution of various failure types in induction motors.

Due to various operational factors in rotating machines, such as mechanical or thermal misalignment, inadequate lubrication, operational fatigue, and environmental corrosion, bearings deteriorate quickly and shut down the processes [8]. If these above mentioned abnormalities are ignored, then rotating machines can develop excessive vibration, overheating, and abnormal noise. Traditional condition-monitoring techniques mainly depend on vibration analysis, and statistical feature extractions have been effective, but they rely heavily on handcrafted features and struggle in noisy environments [9–11]. Moreover, contemporary and current trending monitoring systems generate large, high-dimensional datasets, creating challenges related to feature redundancy, computational load, and the need for fast decision-making. Recent progress in machine learning (ML) and deep learning (DL) has significantly advanced fault diagnosis by enabling models to process large datasets and uncover complex relationships within them [12–14]. These architectures can improve classification performance significantly, but still, they remain sensitive to high-dimensional noisy signals and may fail when multiple fault types coexist simultaneously [15]. To overcome these limitations, researchers increasingly rely on feature-selection strategies and hybrid ML–DL frameworks for detecting and classifying abnormalities in complex environments [16,17]. In this work, we introduce RNN-XBoostNet, a new hybrid approach that uses a Recurrent Neural Network (RNN) to learn temporal patterns from raw time-domain signals and XGBoost to perform robust fault classification in bearing faults classification. The RNN extracts dynamic temporal features of bearing behavior [18], which are then classified by the gradient-boosting-based XGBoost model to identify abnormal machine conditions.

1.1. Challenges in Bearing Fault Detection and Classification

Identifying and classifying abnormal conditions in bearings is an extremely crucial task because these faults often introduce only subtle changes in machine signals, which can easily be masked by noise, load variations, and environmental disturbances. Accurate detection requires high-quality, pre-processed data to separate fault-related patterns from background fluctuations for machines' reliable operation [19,20]. Faulty bearings influence

multiple signal domains, including vibration, acoustic emissions, electrical quantities, and thermal effects, which make their combined interpretation difficult to diagnose accurately. In practical scenarios, fault classes also tend to be highly imbalanced, with some conditions occurring far more frequently than others, leading to biased classifiers and poor performance on rare faults due to the non-stationary and noisy nature of machine signals [21,22]. In addition, industrial diagnostic systems must operate in real time, processing continuous sensor streams with speed, accuracy, robustness, and practical deployment of diagnostic systems [23]. These demands highlight the limitations of traditional signal-processing approaches and motivate the increasing deployment of machine learning and deep learning techniques for more adaptive and robust diagnosis [24,25]. The methods proposed in this study aim to improve sensitivity, enhance classification accuracy, and reduce computational overhead. Table 1 summarizes the main challenges associated with bearing fault detection and classification across different techniques.

Table 1. Challenges in Bearing Fault Detection and Classification.

Challenges	Description
Weak Signal Variations	Faults cause subtle changes, hard to detect under noise, load, and speed shifts.
Complex Signal Representation	Multiple signals (vibration, sound, current, temperature) are affected, making analysis complex.
Class-Imbalance	Some faults dominate data, biasing classifiers and reducing accuracy on rare faults.
Real-Time Constraints	Systems must process data continuously with high efficiency and accuracy.
Advanced Methods	Traditional methods are insufficient; AI approaches improve sensitivity and accuracy.

1.2. Signal Processing Techniques for Fault Detection

In the early stages of bearing fault diagnosis, classical signal-processing techniques such as the Fourier Transform (FT), Wavelet Transform (WT), and Hilbert–Huang Transform (HHT) were widely used for feature extraction [26–28]. These techniques convert time domain signals into the frequency domain, enabling the detection of characteristic harmonics and sidebands, but it performs poorly when signals are non-stationary to identify fault-related patterns. WT addresses this limitation by providing multi-resolution analysis, capturing both short-duration transients and long-term features through variable-speed and load-varying systems. Similarly, HHT effectively handles nonlinear and non-stationary signals by decomposing them into intrinsic mode functions (IMFs) that reflect fault-related patterns and signal-to-noise conditions. However, these traditional methods struggle with noisy, high-dimensional datasets typically produced by modern monitoring systems, prompting a shift toward machine learning (ML) and deep learning (DL) for more advanced real-time diagnostics and applicability [29,30]. Table 2 provides an overview of these signal-processing techniques, highlighting their main strengths, limitations, and applicability to stationary and non-stationary signal analysis.

Table 2. Signal Processing Techniques for Fault Detection.

Technique	Strengths	Limitations	References
FFT	Frequency domain, harmonics, real-time	Stationary signals only	[26]
WT	Multi-resolution, non-stationary, short/long-term	Complex, wavelet-dependent	[27]

Table 2. *Cont.*

Technique	Strengths	Limitations	References
HHT	EMD, IMFs, non-linear/non-stationary	High computation, noise-sensitive, mode mixing	[28]
Traditional Methods	Small dataset feature extraction	Poor on high-dimensional/noisy data	[29,30]

1.3. Machine Learning Approaches for Bearing Fault Classification

The deployment of computer-aided techniques, particularly machine learning (ML), has significantly improved the detection and classification of faults in rotating electrical machines due to their ability to automatically learn complex patterns from extracted features. ML algorithms rely on labeled datasets to learn patterns associated with different fault conditions. Support Vector Machines (SVMs) are commonly used supervised classifiers that separate data using optimal hyper-planes and can be extended to multi-class problems [31]. Random Forests, an ensemble of decision trees, enhance classification accuracy and reduce overfitting through aggregated learning [32]. K-Nearest Neighbor (KNN) classifies samples based on proximity in feature space, although it becomes computationally expensive for large datasets. Decision trees, which split data recursively until terminal nodes represent output classes, form the basis of many ensemble models such as Random Forests and Gradient Boosting [33,34]. Despite their strengths, traditional ML models depend heavily on manual feature extraction. Deep Learning (DL), on the other hand, automatically learns hierarchical representations from data, making it highly effective for high-dimensional, time-series, and end-to-end diagnostic tasks. Convolutional Neural Networks (CNNs), originally developed for image processing, excel in feature extraction and fault identification [35,36]. Recurrent Neural Networks (RNNs) capture sequential dependencies but face vanishing-gradient issues, which are addressed by Long Short-Term Memory (LSTM) networks that preserve long-term temporal relationships, making them well-suited for bearing fault diagnosis [37]. Autoencoders, used in unsupervised settings, detect anomalies by modeling normal machine behavior and identifying deviations. Hybrid ML–DL approaches further enhance performance by combining complementary strengths, for example, LSTM–XGBoost for improved precision, CNN–SVM for stronger feature-based classification, and Random Forest–LSTM for combining dimensionality reduction with temporal modeling [38,39]. Table 3 summarizes key ML and DL methods used in fault diagnosis, highlighting their advantages, limitations, and application areas.

Table 3. Machine Learning and Deep Learning Approaches for Bearing Fault Classification.

Approach/Model	Type	Strengths	Limitations	Use in Fault Diagnosis	References
SVM	Supervised ML	Optimal hyperplane, classification, multi-class	Feature engineering, kernel sensitive, big data issues	Class separation	[31]
RF	Ensemble ML	Robust, regression/classification, feature importance	High cost with many trees, not for sequential	Classification, feature reduction	[32]
KNN	Instance-based	Simple, small data, non-parametric	Heavy for big data, noise sensitive	Baseline classifier	[33]

Table 3. *Cont.*

Approach/Model	Type	Strengths	Limitations	Use in Fault Diagnosis	References
DT	Supervised ML	Simple, interpretable, base for ensembles	Overfitting, less accurate	Used in RF/Boosting	[34]
CNN	Deep Learning	Learns local patterns, raw data, and big data	No long dependencies, high compute	Feature extraction	[35]
RNN	Deep Learning	Sequential/time-series, temporal capture	Suffers from vanishing gradient	Time-series modeling	[36]
LSTM	Deep Learning	handles long-term dependencies, and makes the model more robust	High compute, long training	Temporal features	[37]
AE	Unsupervised DL	Compact reps, anomaly detection	Needs classifier, may miss subtle diff	Unsupervised anomaly	[38]
Hybrid Models	Hybrid ML/DL	Combine strengths, robust, accurate	Complex, high compute	High accuracy	[39,40]

1.4. Current Trends in Bearing Fault Diagnosis

Recent advancements in bearing fault diagnosis have been strongly driven by advances in machine learning, especially deep learning, growing demand for high reliability, early fault detection, and intelligent condition monitoring. Several key trends have emerged that increasingly integrate signal processing, machine learning, and data-driven intelligence to overcome the limitations of traditional methods. Real-time fault detection has become increasingly important with the rise of edge computing and IoT-based monitoring systems, which continuously stream machine-health information for immediate analysis [40]. Ensemble learning has also gained popularity, where multiple models are combined through stacking, boosting, or bagging to achieve higher diagnostic accuracy and robustness [41]. Transfer learning offers another promising direction by enabling models to reuse knowledge from related tasks, reducing the need for large labeled datasets, an important advantage in fault diagnosis, where data collection is often costly and time-consuming [42]. Additionally, as diagnostic models grow more complex, explainability becomes essential. Explainable AI (XAI) techniques help interpret the reasoning behind model decisions, making it easier for engineers to understand why a particular fault type was identified [43]. Table 4 summarizes these emerging trends, including Edge Computing, Ensemble Learning, Transfer Learning, and Explainable AI, which collectively represent the future direction of data-driven fault diagnosis.

Motivated by these limitations, this work proposes a novel CoLaR-FS–RNN-XBoostNet framework, in which a correlation–lasso–recursive feature selection (CoLaR-FS) strategy is tightly integrated with an RNN-based temporal feature extractor and an adaptive XGBoost classifier. Unlike existing hybrid approaches, the proposed framework explicitly emphasizes feature collaboration, noise adaptability, and low-dimensional representation learning, ensuring that only the most discriminative and non-redundant features are forwarded to the classifier. This design not only enhances diagnostic accuracy and robustness but also significantly reduces computational complexity, thereby providing a practical and scalable solution for intelligent bearing fault diagnosis in industrial electrical machines. Table 5

shows the comparative analysis of representative hybrid diagnostic frameworks and the proposed approach for bearing fault diagnostics.

Table 4. Current Trends in Bearing Fault Diagnosis.

Trends	Description	Advantages	Challenges	References
Edge Computing & IoT	Real-time fault detection via on-device and edge-level monitoring.	Continuous monitoring reduces latency better scalability.	Limited computational power and data synchronization issues.	[41]
Ensemble Learning	Combines multiple models (stacking, bagging, boosting) for prediction	Higher accuracy, robustness to noise, and reduced overfitting.	High computational cost, requires careful model selection	[42]
Transfer Learning	Adapts pre-trained models from one domain to another	Less labeled data needed, faster training, better generalization	Domain shift issues, limited adaptability to rare/novel faults.	[43]
Explainable AI (XAI)	Interprets and explains model predictions in diagnosis	Simple to apply	Limited to complex models	[44]

Table 5. Comparative Analysis of Representative Hybrid Diagnostic Frameworks and the Proposed Approach.

Study/Framework	Feature Extraction Strategy	Classifier	Feature Redundancy	Computational Burden	Key Limitations
CNN-based hybrid models	CNN-learned spatial features	SVM/RF	Not addressed	High	Weak temporal modeling
LSTM-only models	LSTM temporal features	Softmax	Not addressed	Moderate–high	Overfitting & Noise sensitive
LSTM–XGBoost [39]	LSTM embeddings	XGBoost	Implicit	Moderate	Feature redundancy
RNN–XGBoost [40]	RNN temporal features	XGBoost	Implicit	Moderate	Limited non-stationary adaptability
Handcrafted + XGBoost	Statistical/spectral	XGBoost	Manual/heuristic	Low–moderate	Expert-dependent
Proposed CoLaR-FS–RNN-XBoostNet	RNN temporal learning	Adaptive XGBoost	Explicit (Correlation + Lasso + RFE)	Low (lightweight)	Threshold sensitivity

Bearing fault diagnostics remains an essential topic in the research area for improving the reliability of rotating electrical machines, and an extensive range of approaches, from classical signal processing to advanced learning-based techniques, continues to be explored. In Ref. [44], the authors proposed wavelet-based features combined with SVM, which achieved high performance even under noisy conditions for bearing fault diagnosis. A CNN–LSTM framework was proposed in Ref. [45] to capture both spatial and temporal characteristics in complex datasets to diagnose bearing faults. In Ref. [46], the authors introduced a hybrid method using Variational Mode Decomposition (VMD) with Random Forests for industrial machines bearing fault classification. In Ref. [47], the authors suggested a Transformer-based architecture to focus on the most informative portions of vibration signals, improving both interpretability and diagnostic accuracy for classifying bearing faults. In Ref. [48], the authors employed Statistical indicators such as RMS,

kurtosis, and skewness were used for detecting abnormal conditions in bearings, while in Ref. [49], the authors utilized frequency-domain characteristics to extract fault-related components in non-stationary environments to diagnose abnormal behaviors in bearings. In Ref. [50], the authors proposed the Short-Time Fourier Transform (STFT) to capture time-varying spectral signatures for detecting abnormal conditions in faulty bearings. In Ref. [51], the authors proposed fault detection performance enhancement by combining empirical mode decomposition (EMD) with deep belief networks (DBNs) for bearing fault diagnosis. In Ref. [52], the authors suggested Attention-based mechanisms in deep learning models, and in Ref. [53], the writers suggested emphasizing key signal features and achieving stronger diagnostic results in classifying abnormal conditions in bearings. In Ref. [54], the authors explored cross-domain ML models that generalize well across datasets with minimal retraining, a valuable property with limited labeled data to diagnose abnormalities in bearings. In Ref. [55], the authors suggested Autoencoders for unsupervised anomaly detection in situations with scarce ground-truth labels, with accuracy parameters to diagnose bearing faults. In Ref. [56], the authors proposed hybrid approaches integrating physical modeling with data-driven analysis to strengthen both robustness and interpretability to detect abnormalities in faulty bearings. In Ref. [57], the authors proposed that ensemble learning combined with PCA was used to reduce computational load while maintaining high accuracy in diagnosing faults in bearings. In Ref. [58], the authors used spectral kurtosis and harmonic amplitudes to enhance frequency-domain diagnostics, while in Ref. [59], the authors relied on wavelet-derived statistical features to improve the classification of abnormalities in bearing faults. In Ref. [60], the authors implemented total harmonic distortion (THD) with power spectral density (PSD) to diagnose bearing faults under varying operating conditions. Table 6 summarizes these studies, highlighting the progression from traditional signal processing toward more capable hybrid and deep learning-based diagnostic techniques.

Table 6. Summary of Literature Review on Bearing Fault Diagnosis.

Reference	Approach	Domain	Pros	Cons	Key Insights
[45]	Wavelet Transform + SVM	Time-Frequency + ML	High diagnostic performance in noisy conditions	Requires careful feature engineering	An effective hybrid of WT feature extraction with an ML classifier
[46]	CNN + LSTM	Deep Learning	Captures spatial & temporal dependencies	High complexity, needs a large dataset	Robust modeling of sequential signals
[47]	VMD + Random Forest	Adaptive Feature Selection + ML	Good for industrial applications	Computationally intensive	Adaptive decomposition improves accuracy
[48]	Transformer-based architecture	Deep Learning	Improved interpretability, attends to important signal parts	Requires significant computational power	Effective for multi-domain databases
[49]	Statistical time-domain features (RMS, Kurtosis, Skewness)	Time-Domain	Simple, interpretable	Not robust to noise, expert-dependent	Traditional baseline approach

Table 6. *Cont.*

Reference	Approach	Domain	Pros	Cons	Key Insights
[50]	STFT	Time-Frequency	Captures time-varying spectral features	Limited resolution trade-off	Improved over FT for non-stationary signals
[51]	Wavelet Transform (WT)	Time-Frequency	Better time-frequency localization	Choice of mother wavelet critical	Effective for transient fault detection
[52]	EMD + DBN	Time-Frequency + Deep Learning	Enhanced fault detection	Sensitive to noise in decomposition	Improves detection with non-linear features
[53]	Attention mechanisms in DL	Deep Learning	Focuses on informative aspects of the signal	Adds computational overhead	Improves interpretability & performance
[54]	Domain adaptation	Transfer Learning	Generalizes across datasets	Needs pre-trained models	Enables cross-dataset robustness
[55]	Unsupervised autoencoders	Unsupervised Learning	Works without labels, anomaly detection	Performance depends on architecture	Useful for unlabeled data
[56]	Fusion of physical modeling + data analysis	Hybrid	Incorporates domain knowledge	Complex integration	Improves robustness and physical interpretability
[57]	Lightweight neural networks	Edge/Embedded DL	Efficient for edge devices	Lower representational capacity	Optimized for real-time edge computing
[58]	Ensemble learning + PCA	ML + Dimensionality Reduction	Reduces computational load, high accuracy	Interpretability may be reduced	Balances efficiency and accuracy
[59]	Spectral Kurtosis + Harmonics	Frequency-Domain	Provides fault-sensitive features	Feature engineering needed	Enhances frequency domain representation
[60]	THD + PSD	Frequency-Domain	Effective under different conditions	Requires careful preprocessing	Key harmonic and power spectral features

The main objectives and contributions of this article are summarized below:

- Introduced RNN-XBoostNet, a hybrid diagnostic framework that combines RNNs for time-series feature extraction with XGBoost for robust classification of bearing faults in rotating electrical machines.
- Introduced CoLaR-FS, a novel feature selection method that integrates correlation, Lasso regression, and RFE for dimensionality reduction and selection of the most discriminative fault features.
- Established an RNN–XGBoost synergy to model complex and nonlinear fault patterns using a combination of temporal and learned features.
- Validated the framework with Tallinn University of Technology and Induction Machine (IM) datasets, which comprised ten fault type instances under realistic operating conditions.

- Conducted a comprehensive evaluation on the CWRU and IM datasets to assess model generalization, robustness, and computational efficiency.
- The integration of CoLaR-FS with RNN-XBoostNet led to substantial accuracy gains, improving performance from 87.01% to 98.35% on CWRU and from 79.98% to 98.57% on the IM dataset.

2. CoLaR-FS (Proposed Feature Selection Mechanism)

CoLaR-FS (Correlation-based Lasso Regression with Recursive Feature Elimination) is a structured hybrid feature selection approach to enhance both the accuracy and computational efficacy of bearing fault diagnostics. It is a feature selection technique with a comprehensive approach leveraging three synergistic components, Correlation Filtering, Lasso Regression, and Recursive Feature Elimination (RFE), concluding with classification via the hybrid LSTMBoostNet.

2.1. Correlation-Based Filtering

The objective of this stage is to decrease feature redundancies by mitigating overfitting from feature pairs with high Pearson correlation coefficients $(|p_ij| > \tau)$ by removing one of the features from each pair. This guarantees that only independent, informative features remain for subsequent processing [61].

2.2. Lasso Regression

The Lasso (Least Absolute Shrinkage and Selection Operator) technique incorporates $L1$ regularization to impose penalties on the coefficients of the features, resulting in the near-zero coefficients of the relatively less informative features and exclusion of those features. This process improves both the sparsity and interpretability of the model [62].

2.3. Recursive Feature Elimination (RFE)

During each of the iterations that the model is trained, RFE systematically eliminates the features judged as less important according to the feature importance rankings until the optimal subset is reached. This process ensures that only the most discriminative features remain for classification. The LSTM component of the RNNBoostNet then analyzes the selected feature set to identify the temporal dependencies that exist in the signal patterns (e.g., vibration or current), and the final fault classification is performed by the XGBoost classifier with high accuracy and computational efficiency. The combination of these processes establishes a robust and clear diagnostic procedure for bearing faults [63,64]. Figure 2 shows the working of CoLaR-FS, which is employed for the selection of optimal features from the dataset, while Algorithm 1 presents the detailed algorithm of CoLaR-FS (Correlation–Lasso–RFE combined with the RNN-XGBoost feature selection process).

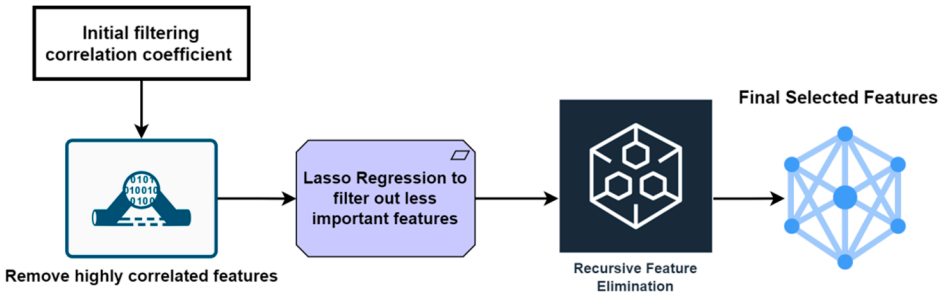


Figure 2. Proposed CoLaR-FS Framework for Optimal Feature Selection from the Dataset.

2.4. Mathematical Background of CoLaR-FS

Let the input and output features from the dataset be given as:

$$X = [x_1, x_2, \ldots\ldots, x_n] \in R^{m*n}, y \in R^m \tag{1}$$

in which m is the number of samples, n Number of initial features, $x_i \in R^m$ is the ith feature vector, and y is the target vector representing fault labels. Eliminate features with low linear association to the target variable and reduce high inter-feature correlation to remove highly correlated features to avoid redundancy. For each pair of features X_i and X_j we calculate the Pearson correlation coefficient p_{ij} as:

$$p_{ij} = \frac{Cov(X_i, X_j)}{\sigma X_i, \sigma X_j} \tag{2}$$

where, $Cov(X_i, X_j)$ is the covariance between features X_i and X_j, and σX_i and σX_j are the standard deviations of X_i and X_j for a threshold τ. If $|\, p_{ij} \,| > \tau$, remove one of the features X_i or X_j to remove features weakly correlated with the target, and perform feature selection by applying Lasso (Least Absolute Shrinkage and Selection Operator) Regression to filter out less important features. Mathematically:

$$minimize\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j X_{ij})\right) + \lambda\sum_{j=1}^{p}|\beta_j| \tag{3}$$

in which y_i is the target variable, X_{ij} is the value of the jth feature for the ith observation, β_j is the coefficient of the jth feature, λ is the regularization parameter controlling the sparsity of the model, and $\beta_j = 0$ denotes the excluded Features. Refine the feature set by recursively removing the least important features. The steps involved in RFE are given as:

- Let $f : R^{m*n}$ be a base estimator.
- Train f on the current p-dimensional feature set.
- Obtain feature importance scores $I = [I_1, I_{2,\ldots,}, I_k]$
- Remove the feature with the smallest I_k.
- Repeat until $k \ll n$ features remain.

The final selected feature matrix is given as:

$$X_{CoLaR=}[x_{i1}, x_{i2}, \ldots\ldots, x_{ik}] \tag{4}$$

After obtaining the optimized feature set, X_{CoLaR}, a hybrid model is employed to capture temporal dependencies and perform robust fault classification. The selected features are reshaped into sequences to capture temporal patterns in signals such as vibration or current. The RNN updates hidden states as follows:

$$h_t = f(x_t, h_{t-1} : \theta) \tag{5}$$

in which h_t is the hidden state at time t, x_t is the input at time t, and θ represents the trainable parameters for the final output h_{final} that encodes the sequential characteristics of the fault signals. The RNN features h_{final} are input to XGBoost, which builds an ensemble of boosted decision trees. For iteration t, the prediction is:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i), \qquad f_k \epsilon F \tag{6}$$

in which F shows the space of regression trees and f_k is the kth tree that XGBoost minimizes the regularized objective function given as:

$$\beth(\theta) = \sum_{i=1}^{m} l(y_i, \hat{y}_i) + \sum_{k=1}^{t} \Omega(f_k) \tag{7}$$

Then RNN outputs to the XGBoost model that is trained on the LSTM outputs and produces the final prediction that is y_{final}. Mathematically:

$$y_{final} = arg \max_{c} P\left(y = c \middle| h_{final}\right) \tag{8}$$

in which l the loss function and Ω is the regularization term and y_{final} shows the final predicted features.

$$X \overset{CoLaR-FS}{\rightarrow} X_{CoLaR} \overset{RNN}{\rightarrow} h_{final} \overset{XGBoost}{\rightarrow} y_{final} \tag{9}$$

Algorithm 1: CoLaR-FS (Correlation–Lasso–RFE + RNN-XGBoost Feature Selection)

Input: Dataset D with feature matrix $X(n * p)$, Target variable y
Output: Final Selected optimal Feature subset X'
STEP 1: Correlation-Based Redundancy Removal
Initialize feature set $X' = X$, for each pair of features $(X_i$ and $X_j)$ in X' Initialize threshold τ (e.g., τ = 0.9)
Calculate the Pearson Correlation Coefficient using p_{ij}

$$p_{ij} = \frac{Cov(X_i, X_j)}{\sigma X_i, \sigma X_j} \tag{10}$$

If $\left|p_{ij}\right| > \tau$: ($\tau$ is the correlation threshold) remove one of the features X_i and X_j based on lower mutual information with y
STEP 2: Lasso Regression for Sparsity-Based Filtering
Fit Lasso Regression on the filtered dataset X' with target variable y
For each feature j in X':

$$minimize\left(\frac{1}{2n}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\beta_j X_{ij})\right) + \lambda\sum_{j=1}^{p}|\beta_j| \tag{11}$$

If $\beta_j = 0$: (β_j is the Lasso coefficient for feature j) then add feature j to the new feature set X' and update feature set
STEP 3: Recursive Feature Elimination (RFE)
Initialize the desired number of features k, while $|F| > k$
Train an estimator using the RNN-XBoostNet and Rank features based on their importance (e.g., the magnitude of coefficients $\left|\beta_j\right|$)
Remove the least important feature from S and $X_{RFE} =$ be the refined feature subset
STEP 4: LSTM Feature Extraction
Train RNN (LSTM layers on X_{RFE}) to capture temporal dependencies:

$$h_t = f\left(W_{ihx_t} h_{t-1} + W_{hh} h_{t-1} + b_h\right) \tag{12}$$

$Z = [z_1, z_2, \ldots, z_n]$ = be the feature representations from LSTM

Algorithm 1: *Cont.*

STEP 5: Classification Using RNN-XBoostNet
Prepare input data for RNN-XBoostNet using the selected feature set S
Set the weight matrices and biases for the RNN.
Train the RNN-XBoostNet model on the selected features S with target y
Rank features in X_{RFE} using XGBoost feature importance score (Gain or SHAP).

$$y_{final} = \underbrace{min}_{\theta} \sum_{i=1}^{l} l(y_i, \hat{y}_i) + \Omega(\theta) \quad (13)$$

STEP 6: Final Feature selection subset
Select the top-k most informative features based on the combined LSTM–XGBoost score.
Return : Final selected feature subset $F_optimal$ used by LSTM + LightGBM
final optimal feature subset

$$X^* \subseteq X_{RFE} \quad (14)$$

2.5. Parameter Sensitivity Analysis of CoLaR-FS

The proposed feature selection mechanism, named CoLaR-FS (Correlation–Lasso–Recursive Feature Selection), integrates multiple selection stages to balance feature relevance, redundancy reduction, and classification performance. In the correlation-based filtering stage, a redundancy threshold (τ) is used to eliminate highly correlated features, as τ was varied across 0.80, 0.85, 0.90, and 0.95, while all other parameters were kept constant. For each value, the resulting number of selected features and the corresponding classification accuracy were evaluated using the proposed RNN-XBoostNet framework. In this study, λ was optimized using k-fold cross-validation, where multiple candidate values were evaluated to identify the configuration that maximized validation accuracy while maintaining feature stability. The analysis shows that overly large λ values eliminate informative features and degrade classification performance, whereas excessively small values fail to suppress weak and noisy features. The selected λ value provides stable performance across both datasets, confirming the robustness of the Lasso-based selection stage within the CoLaR-FS framework, as shown in Table 7.

Table 7. Influence of correlation threshold τ on feature count and classification accuracy for CoLaR-FS framework.

Correlation Threshold (τ)	Number of Selected Features	Classification Accuracy (%)
0.80	24	94.12
0.85	31	96.47
0.90	38	98.57
0.95	52	98.11

3. Preliminaries

3.1. Stacked Recurrent Neural Network (RNN):

A stacked recurrent neural network (RNN) is an advancement of basic RNN architecture where several RNN layers are stacked to learn hierarchical temporal dependencies present within certain sequences of data [65]. In this architecture, each hidden state output of a lower layer becomes the input to the layer above it. For example, consider an input

sequence $X = \{x_1, x_2, \ldots x_T\}$ with $x_t \epsilon R^d$ the hidden state of the first RNN layer at time step t is computed as:

$$h_t^{(1)} = f^{(1)}\left(x_t, h_{t-1}^{(1)}; \theta^{(1)}\right) \tag{15}$$

in which $f^{(1)}$ shows the recurrent function, $h_t^{(1)}$ is the hidden state and $\theta^{(1)}$ denotes the parameters of the first layer. For subsequent layers, the hidden state is updated using the output of the previous layer as input, such that:

$$h_t^{(l)} = f^{(l)}\left(h_{t-1}^{(l)}, h_{t-1}^{(l)}; \theta^{(l)}\right) l = 2, \ldots, L, \tag{16}$$

in which L is the total number of stacked layers. Finally, the top layer produces a hidden representation $h_t^{(L)}$ and for classification tasks, the final prediction can be derived from the last hidden state through an output mapping function g expressed as:

$$\hat{y} = g\left(h_t^{(L)}; \theta^o\right) \tag{17}$$

This hierarchical formulation allows stacked RNNs to learn both low-level and high-level temporal features, thereby improving their ability to model complex sequential patterns compared to a single-layer RNN [66]. Figure 3 shows the Stacked Recurrent Neural Network (RNN) Architecture for Sequential Data Processing to obtain feature extraction and selection from the most important dataset features.

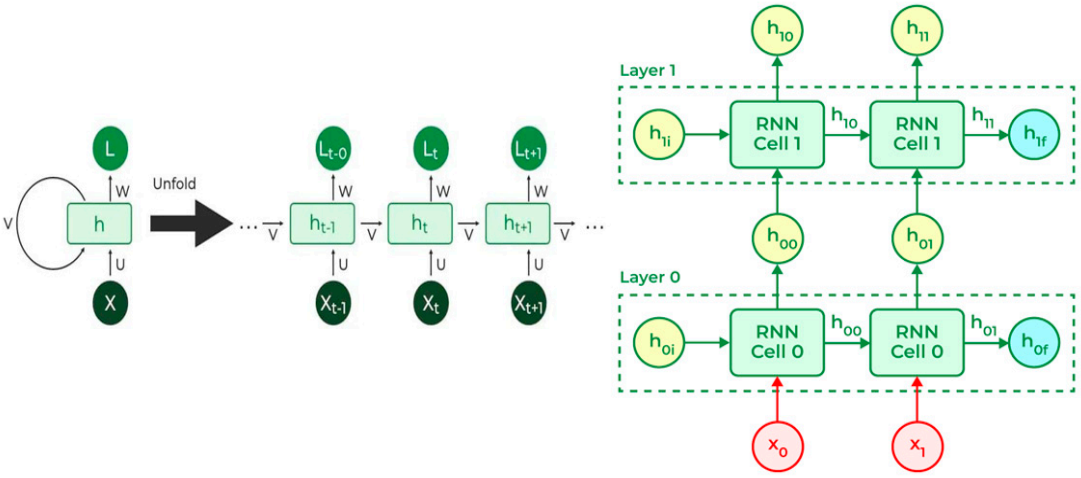


Figure 3. Stacked−Recurrent Neural Network (RNN) Architecture for Sequential Data Processing.

3.2. Extreme Gradient Boosting (XGBoost) Algorithm

XGBoost (Extreme Gradient Boosting) is a fast and large-scale tree-boosting system, which is a sequential learning method. XGBoost greedily creates decision trees, and each decision tree aims to learn from the mistakes made in the previous trees. It is used extensively due to its high performance, and it performs well when the data has many dependencies. XGBoost is developed based on a boosting framework under the assumption that a set of weak learners could be combined to form a strong learner [67]. There is L1 (Lasso) and L2 (Ridge) regularization included in XGBoost, which greatly mitigates overfitting while enhancing overall performance. The main property of XGBoost is the possibility to use parallel computing when learning and building decision trees, in particular. Since XGBoost does not support missing values, it is suitable for datasets that do not contain many missing values or require the use of synthetic data to be incorporated. XGBoost builds the trees in a depth-first manner and uses the technique "max_depth" to prune the trees, to avoid overfitting [68]. In XGBoost, the loss function is also called binary cross-entropy loss,

which is the regularization term for the predicted probabilities p_i and the true labels y_i. Mathematically:

$$L = -\frac{1}{2}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{18}$$

In the above equation, N is the total number of samples in the dataset, y_i actual values for the ith sample p_i the predicted probability of a positive class $\log(p)$ is the natural logarithm function. XGBoost is transformed into a GBM when γ and λ are set to 0. The regularization term added loss function as in the given equation.

$$Loss = L + \Omega(f) \tag{19}$$

where $\Omega(f)$ is the regularization term for the model, which can be expressed as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{t=1}^{T} | \, | \theta_t | \, |^2 \tag{20}$$

In the above equation, γ is the regularization used to penalize the no. of trees λ is the regularization term, weighted square of trees, T stands for the total no. of trees, and θ_t Are the weights assigned to the t–th tree. The structure of the XGBoost model used in our proposed methodology is illustrated in Figure 4, showcasing its gradient-boosting mechanism for effective classification.

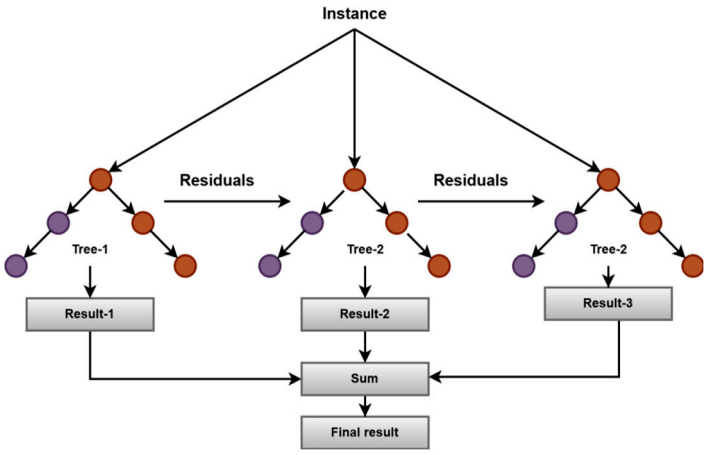


Figure 4. XGBoost Model Architecture for Efficient Classification.

3.3. Mathematical Framework of the Proposed Framework RNN-XGBoost for Bearing Fault Classification

The two developing approaches have been combined into the RNN-XGBoost, which uses Recurrent Neural Networks (RNN) for learning sequences from time-series data and uses Extreme Gradient Boosting (XGBoost) for accurate classification. In this paper, the mathematical formulation is discussed in detail, given that the primary problem is a classification problem. The first approach is data preprocessing: given a dataset $X \in R^{n*d}$ where n is the number of samples and d is the number of features. Each feature is normalized by applying mean ($\pi = 0$) and standard deviation ($\sigma = 1$) to ensure numerical stability. Mathematically:

$$X' = \frac{X - \mu}{\sigma} \tag{21}$$

The CoLaR-FS technique is applied to extract the most discriminative features for fault diagnosis. RNN allows for sequence information to be stored, learned, and used in the analysis of time-series inputs. Input Sequence: X_t at each time step t, and the hidden state

h_t at each time step t is updated recursively based on the previous hidden state h_{t-1} and the current input as x_t. The RNN computes the hidden state h_{t-1} at time step t using the following update rule:

$$h_t = \sigma(W_{hh}h_{t-1} + W_{xh}X_t + b_h) \tag{22}$$

where W_{hh} and W_{xh} are weight metrics, b_h is the bias term, and σ is the activation term. The final output of the RNN as h_T, where T is the last time step, is passed through a dense layer to produce the feature representation of the sequence:

$$\hat{X} = W_{hy}h_T + b_y \tag{23}$$

where $\hat{X} \in R^m$ Is the vector of extracted features from CoLaR-FS for the next classification step. The output from the RNN is $\hat{X}$ that is fed to the XGBoost classifier takes its input from the efficient gradient boosting, which constructs the model through decision trees. We will use XGBoost to train the data because, during learning, it seeks to minimize the following objective function.

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \tag{24}$$

where $l(y_i,\ \hat{y}_i)$ is the loss function, typically the log-loss for multi-class classification, $\hat{y}_i$ is the predicted probability for class i, f_k is the k-th tree in the ensemble, and $\Omega(f_k)$ is the regularization term to prevent overfitting. XGBoost applies regularization to control the complexity of the model:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\sum_{j=1}^{T} \omega^2{}_j \tag{25}$$

where T shows the total number of leaves in the decision tree, γ and λ are the regularization parameters, and $\omega^2{}_j$ shows the weights assigned to the leaf. XGBoost grows trees sequentially in the following manner by adding a new tree node that splits depending on the decision of the gradient of the loss function. At each iteration t, the prediction is updated using the gradient of the loss function.

$$\hat{y}_t = \hat{y}_{t-1} - \eta.grad_t \tag{26}$$

where η is the learning rate, and $grad_t$ is the gradient of the loss concerning the output of the t-th tree. After training, the RNN produces sequential features from new bearing fault data for classification using the XGBoost model. The final output probability for each fault class is calculated using the softmax function:

$$\hat{p}_i = \frac{\exp(f_k(X))}{\sum_{j=1}^{C} \exp(f_k(X))} \tag{27}$$

where $\hat{p}_i$ stands for probability, $f_k(X)$ is the output of the k-th model of the XGBoost tree, and is the fault types that are predicted by the proposed architecture.

3.4. Hyperparameter Searchfor the Proposed Architecture RNN-XBoostNet

The effect that hyperparameter values have on the efficiency of the performance of machine learning algorithms is important. Regarding the RNN, the hyperparameters of greatest importance are the number of layers (the depth of the model), hidden units (the learning capacity), learning rate (size of the weight update), and dropout rate (overfitting regulation). Important hyperparameters for XGBoost are the number of estimators (the number of trees), maximum depth (tree complexity), learning rate (size of the update step), and the subsample ratio (which prevents overfitting). GridSearchCV is employed

to determine this combination to systematically evaluate various hyperparameters to best optimize the learning and generalization performance [69]. Table 8 shows the Optimal Hyperparameter Configuration of the RNN-XGBoost Hybrid Model for Fault Diagnosis.

Table 8. Optimal Settings of the Hyperparameters for the RNN-XGBoost Hybrid Model for Fault Diagnosis.

Hyperparameter	Description	Optimal Value	Component
Number of Layers	Defines the depth of the RNN model; controls sequential learning capacity.	2	RNN
Hidden Units (per layer)	Determines the number of neurons per RNN layer; affects model complexity.	128	RNN
Dropout Rate	Prevents overfitting by randomly dropping neurons during training.	0.3	RNN
Learning Rate (Adam)	Controls the step size for weight updates using the Adam optimizer.	0.001	RNN
Batch Size	Number of samples processed before model weight updates.	64	RNN
Epochs	Number of complete passes through the training dataset.	100	RNN
Number of Estimators	Total number of boosting rounds in XGBoost.	500	XGBoost
Max Depth	The maximum depth of each decision tree prevents overfitting.	6	XGBoost
Learning Rate (η)	Shrinkage parameter that controls learning speed and regularization.	0.1	XGBoost
Subsample	Fraction of samples used for fitting individual trees to prevent overfitting.	0.8	XGBoost
Colsample_bytree	Fraction of features used per tree; improves generalization.	0.8	XGBoost
Regularization (λ)	L2 regularization term to reduce model complexity.	1	XGBoost
Gamma (min_split_loss)	Minimum loss reduction required to make a further partition on a leaf node.	0.2	XGBoost

4. Validation of Proposed CoLaR-FS-(RNN-XGBoost)

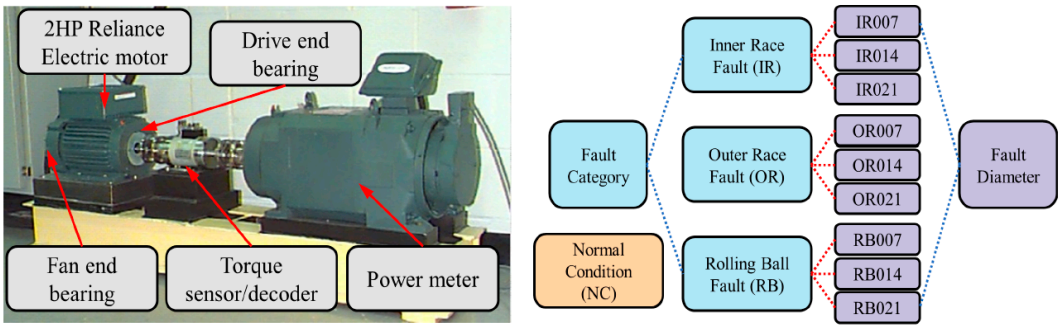
To validate the proposed CoLaR-FS–RNN-XBoostNet framework under both benchmark and realistic industrial conditions, experiments were conducted using a laboratory-scale induction machine (IM) test rig designed for bearing fault diagnosis. The experimental setup consists of a three-phase squirrel-cage induction motor mechanically coupled to a load machine through a flexible coupling. The Case Western Reserve University (CWRU) dataset is based exclusively on vibration signal measurements acquired using accelerometers mounted on the bearing housing. In this study, vibration signals from the CWRU dataset are treated as independent time-series inputs and processed using the same feature extraction, selection, and classification pipeline as other diagnostic signals. It is important to emphasize that vibration signals are not combined with electrical or mechanical signals. Instead, the CWRU dataset is used as a benchmark dataset to evaluate the generalization capability of the proposed CoLaR-FS–RNN-XBoostNet framework across different sensing modalities. This signal-agnostic design allows the proposed approach to be applied consistently to vibration-based as well as current-based diagnostic scenarios.

4.1. Case I: Case Western Reserve University (CWRU) Bearing Diagnosis Dataset

The Bearings Data Center, Case Western Reserve (CWRU) University dataset remains pertinent for bearing fault diagnosis Literatureand its statistics areas shown in Table 9. CWRU also tailored its dataset to mirror authentic industrial circumstances, embedding fault-bearing components: inner race, outer race, and rolling elements, to manufactured faults interleaved with electrical discharge machining (EDM) fault sizes; 0.007″ & 0.040″. The dataset also contains an experimental paradigm made up of a 2 HP induction motor, shaft, bearings, and a dynamometer [70]. Accelerometers were used to capture deflection data streams at differing points, which were the drive end (DE) and the fan end (FE) of the motor and the base, while alternating load conditions (0–3 HP) and sample rates (12 kHz and 48 kHz) were in place. The dataset includes a diverse set of fault types, severities, and operating conditions, allowing for thorough analysis of fault detection, classification, and severity levels. The wide variety of fault types and sizes, as well as operating conditions present in the dataset, allows for a comprehensive assessment of fault detection, classification, and severity analysis [71]. The components outlined in Figure 5 illustrate the arrangement of the CWRU test rig motor, dynamometer, torque transducer, and the data acquisition system utilized for the collection of vibration signals while varying the fault conditions present in the bearings.

Table 9. Statistics of the CWRU dataset.

Sr.No	Fault Type	Sampling Frequency/KHz	Fault Diameters (Inches)	Fault Labels
1	Healthy	48	0	Normal
2	Inner Race	48	0.007	IR007
3	Inner Race	48	0.014	IR014
4	Inner Race	48	0.021	IR021
5	Outer Race	48	0.007	OR007
6	Outer Race	48	0.014	OR014
7	Outer Race	48	0.021	OR021
8	Ball Bearing	48	0.007	BB007
9	Ball Bearing	48	0.0014	BB014
10	Ball Bearing	48	0.0021	BB021



(a) (b)

Figure 5. CWRU dataset introduction (**a**) experimental device (**b**) bearing fault modes [72].

4.2. Case II: Induction Machine Dataset of Tallinn University for Bearing Faults Diagnosis

The experimental setup features two identical shaft-coupled induction machines, where one works as a load while the other one is modified to have different bearing faults. 10 fault conditions were examined, which included inner race, outer race, combined inner and outer race, overheating, corrosion, and separator fault, lack of lubrication, and artificially induced 1A and 10A electrical current faults over a period of 7 days, as well as a healthy condition for baseline comparison. A Dewetron data acquisition system recorded signals of current, voltage, speed, and torque while a 20 kHz sampling rate was maintained for all parameters. For accurate measurements of vibrations, 3 accelerometers were 120 degrees apart. Both machines were driven under full load conditions using ABB industrial drives (ACS600 and ACS800), where scalar and direct torque control (DTC) modes were applied. Overall, the dataset consists of 10 faulty bearing conditions, where each condition hosts over 4.1 million data points or about 413,711 samples for each condition. For model development, 70% of the data is allocated for training the model, while the remaining 30% is reserved for model testing. Figure 6 illustrates the experimental setup of the IM bearing fault dataset with induction motor parameters. Table 10 highlights the description of the induction machine (IM) bearing fault dataset and corresponding labels.

Parameter	Value
Number of poles	4
Number of phases	3
Connection	Star/Delta
Number of stator slots	36; non-skewed
Number of rotor slots	28; skewed
Terminal voltage	690V/400V @50 Hz
Rated current	8.8A/13.5A
Rated power	7.5 kW@50 Hz

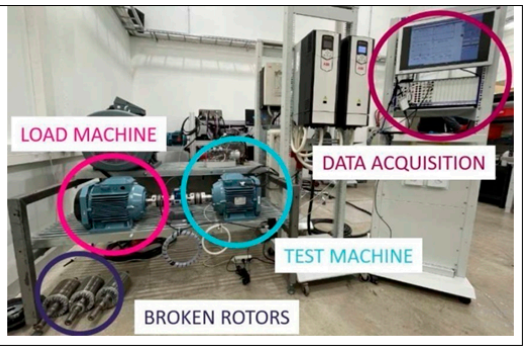


Figure 6. Experimental setup and motor parameters.

Table 10. Description of Induction Machine (IM) Bearing Fault Dataset and Corresponding Labels.

Sr.No	Fault Category	Description	Sampling Frequency (KHz)	Fault Labels
1	Healthy	Normal running condition	20	Normal
2	Inner race	Localized defects on the inner race	20	IR
3	Outer race	Localized defects on the outer race	20	OR
4	Inner race + outer race	Damaged both inner and outer races	20	IR + OR
5	Overheated	Fault caused by thermal stress	20	OH
6	Corroded	Fault due to corrosion/moisture	20	Cd
7	Separator fault	Damaged due to the rolling cage element	20	SF
8	Without lubricant	Absence of lubrication	20	WL
9	1A-7days	Fault induced by 1A supply	20	1A
10	10A-7days	Fault induced by 7A supply	20	10A

The frequency-domain features extracted from the induction motor signals contain important information regarding the machine's operational state and fault behavior. These features describe the variations associated with energy distribution, complexity of the signal, and the harmonic content of the signal, all important for distinguishing healthy and

faulty conditions. Analysis of the magnitude and the dispersion of the spectral components identifies the subtle differences in the frequency response patterns corresponding to the fault type. The statistical and spectral features of the transformed signals provide a complete description of the mechanical and electrical anomalies of the system, thus increasing the fault detection robustness and sensitivity. The features extracted provide a complete description of the dynamic behavior of the system, which can be used for advanced diagnostic models. Table 11 shows the extracted statistical and frequency-domain features for fault diagnosis in induction motors.

Table 11. Feature Derivation and Analytical Expressions for Fault Characterization.

Feature ID	Features Name	Mathematical Relation	Diagnostic Meaning
F1	Mean (μ)	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$	Represents the average value of the signal, baseline level of vibration, or current
F2	Variance (Var)	$Var = \sigma^2$	Quantifies overall variability in signal amplitude
F3	Skewness (SK)	$SK = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \pi}{\sigma}\right)^2$	Indicates symmetry of the signal; non-zero skewness shows fault-induced distortion
F4	Signal Energy (E)	$E = \sum_{i=1}^{N} {x_i}^2$	The total energy content of the signal increases with fault severity.
F5	Amplitude at 50 Hz (A_{50})	Amplitude of FFT at fundamental frequency (50 Hz)	Represents the fundamental component of supply
F6	Peak Frequency (f_{peak})	$f_{peak} = arg\max_f \lvert X(f)\rvert$	Peak frequency represents the frequency component at which the signal's magnitude is maximum.
F7	Bandwidth (BW)	$BW = \sqrt{\frac{\sum (f_i - f_c)^2 . \lvert X(f_i)\rvert}{\sum \lvert X(f_i)\rvert}}$	Bandwidth measures the spread of frequencies around the center frequency.
F8	Kurtosis	$K = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3$	Kurtosis measures signal impulsiveness to indicate high localized defects.

4.3. Proposed Methodology

The proposed methodology for bearing fault diagnosis follows a practical and structured workflow as shown in Figure 7. The process starts from data acquisition, where related signals are first collected under full-load conditions from current, voltage, torque, and speed signals to capture realistic machine behavior. After acquisition, the signals undergo preprocessing to remove noise, followed by feature selection using the CoLaR-FS method, which reduces redundancy and improves model efficiency. Two datasets are used for validation: the benchmark CWRU dataset and the experimentally collected IM dataset, ensuring both theoretical and real-world relevance. Hyperparameters are optimized using GridSearchCV, and the selected features are used to train the hybrid RNN-XGBoost model, which combines temporal learning with robust classification. The trained model is then evaluated on both datasets, demonstrating strong performance on established benchmarks and practical operating conditions.

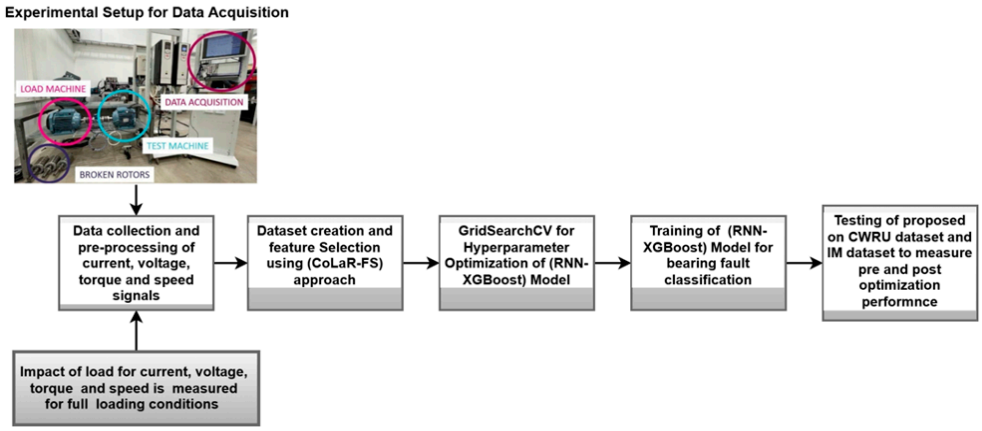


Figure 7. Proposed flow diagram based on (RNN-XGBoost) to diagnose abnormal conditions in bearings.

5. Results and Discussions

The experimental results demonstrate that the proposed CoLaR-FS–RNN-XBoostNet framework consistently outperforms conventional and hybrid diagnostic models across multiple datasets and fault categories. The effectiveness of the proposed RNN-XBoostNet framework largely relies on the implementation of CoLaR-FS, a newly developed feature selection approach that aims at the selection of features that are informative while excluding the out-of-formative features. CoLaR-FS follows sub-ranking, where it first uses correlation analysis, then uses Lasso regression, and finally recursive feature elimination (RFE) to strike a balance between less dimensionality and high relevance of faults [72]. This process not only increases the speed of computation of the model but also increases diagnostic accuracy by paying particular attention to features that are directly relevant to the identification of faults and their classification. In the context of bearing fault diagnosis, CoLaR-FS screened out features from both the time domain and frequency domain of the signals. These features include moments of the current, voltage, torque, and speed signals, their mean-variance and skewness, and kurtosis coefficients in the frequency domain. Figure 8 shows the Feature Correlation Heat map (Selected by CoLaR-FS), which shows the correlation coefficients between motor features such as currents (Ia, Ib, Ic), voltages (Va, Vb, Vc), Torque, and Speed.

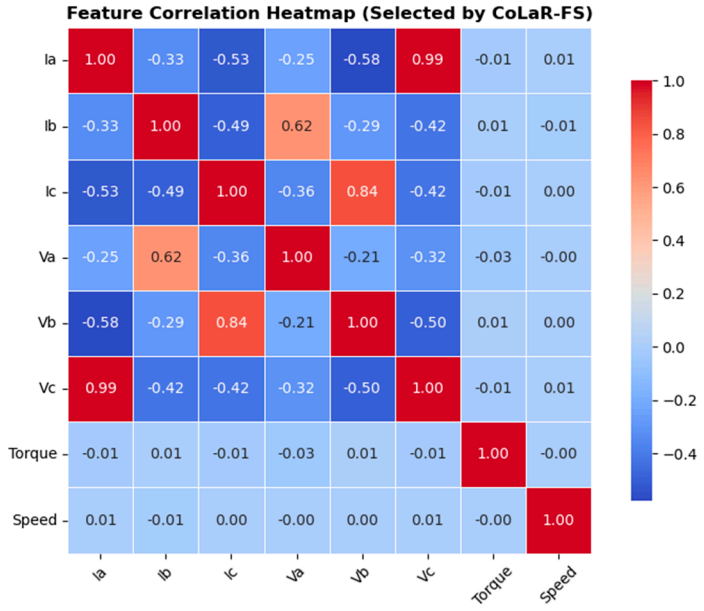


Figure 8. Feature–Correlation Heat map (Selected by CoLaR-FS).

5.1. Characterizing Bearing Faults Through FFT Spectrum Analysis in Electrical Machines

To analyze the frequency-domain characteristics of bearing faults in the induction machine, log-scale Fast Fourier Transform (FFT) spectrums (in dB) were computed for the current, voltage, torque, and speed signals employing bearing health conditions. In this case, Fast Fourier Transform (FFT) analysis was performed on current, voltage, torque, and speed signals of an induction machine operating under ten bearing conditions to examine the fault-related frequency characteristics. The analysis focused on spectral components at the fundamental supply frequency of 50 Hz and odd harmonics. In the healthy condition, the spectrum shows a single dominant peak at 50 Hz with minimal harmonic activity, indicating stable rotor–stator interaction. In faulty bearing cases, additional sideband frequencies and harmonic components emerge around the fundamental and its multiples. These frequency components become more pronounced for severe faults such as inner-race, outer-race, and combined race defects, reflecting the modulation of stator currents due to mechanical disturbances and air-gap irregularities. Log-Scale FFT of voltage signals' spectra exhibits comparatively stable fundamental frequency behavior under healthy operation. However, faulty bearing conditions introduce subtle harmonic variations and spectral spreading, particularly at higher frequencies. Although voltage signals are less sensitive than current signals, their frequency-domain variations provide supplementary diagnostic information associated with electromagnetic coupling and load fluctuations. Figure 9 shows the Log-scale FFT spectra (in dB) of current, voltage, torque, and speed signals, illustrating the frequency-domain characteristics of all ten bearing fault types in the induction machine dataset.

To investigate the non-stationary characteristics of bearing faults in the induction machine, the Continuous Wavelet Transform (CWT) was applied to the current, voltage, torque, and speed signals. CWT provides a joint time–frequency representation, enabling the identification of transient events, frequency modulations, and fault-induced disturbances that are not clearly observable in the time or frequency domains alone. The CWT scalogram of current, voltage, torque, and speed signals provides clear distinctions between healthy and faulty bearing conditions. Under healthy operation, the energy is predominantly distributed and concentrated around the fundamental frequency at 50 Hz with minimal time–frequency dispersion, and these disturbances reflect speed oscillations and slip variations induced by bearing defects, especially under severe fault conditions. The differences in the energy distribution patterns, as illustrated in Figure 10, highlight the capability of CWT scalograms to capture fault-related signatures and differentiate between healthy and defective operating states effectively.

To gain a deeper understanding of the distribution and relationships within the datasets, the dimension reduction technique named t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm was employed. As a nonlinear dimensionality reduction technique effectively maps the high-dimensional feature spaces into two- or three-dimensional representations, allowing visual interpretation of intrinsic data structures such as clusters, separability, and continuity [73]. In this study, t-SNE was applied to both the CWRU dataset and our custom induction motor (IM) dataset to provide a clear and intuitive assessment of the degree to which different feature groups were distinct or overlapping based on their underlying correlations, as shown in Figures 11 and 12. This visualization step enabled an intuitive assessment of how distinct or overlapping various feature groups were based on their underlying correlations. Moreover, the t-SNE plots provided valuable insights into the intrinsic dimensionality of the datasets, facilitating a more informed application of the subsequent feature selection and classification stages using the RNN–XGBoost framework.

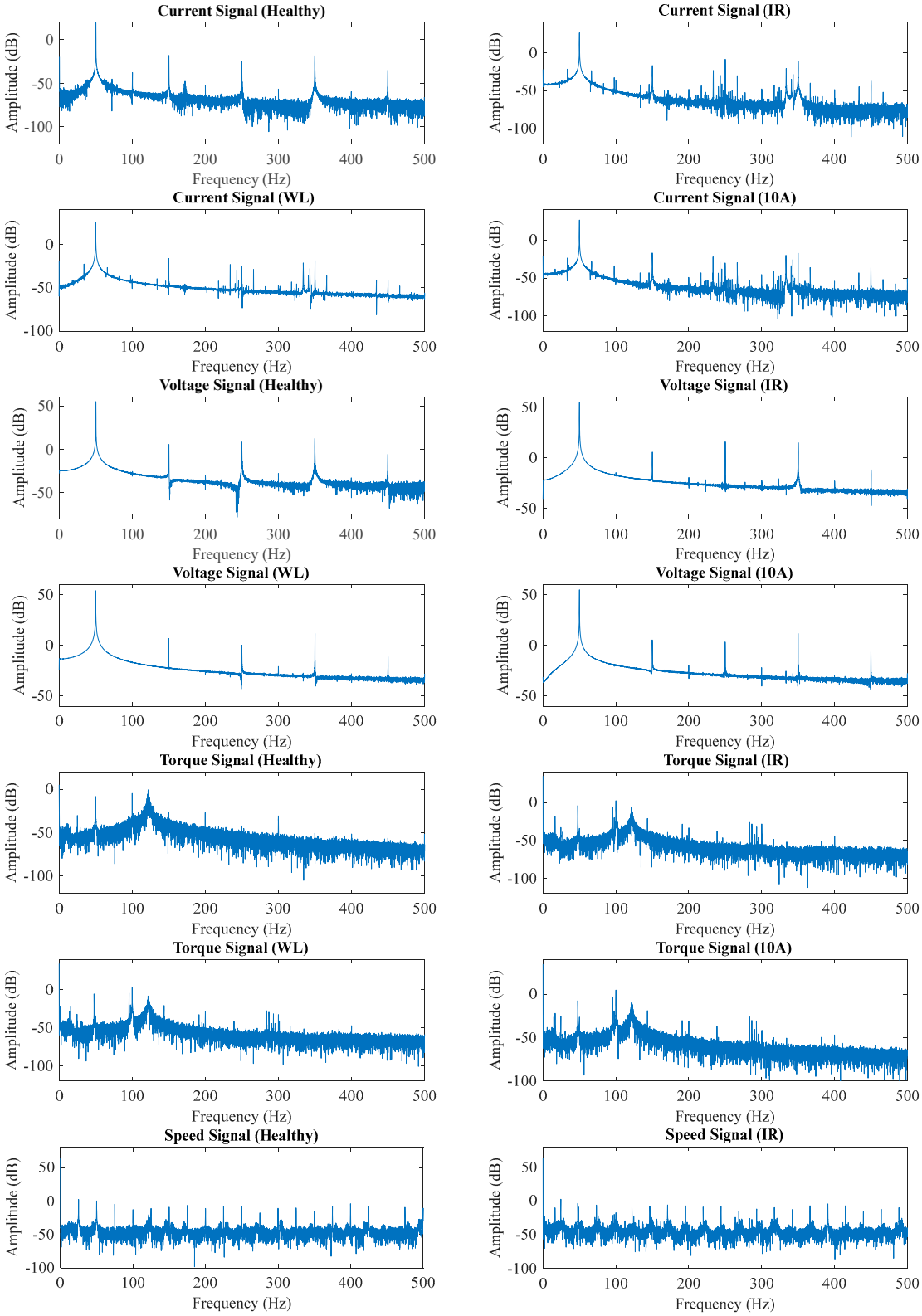


Figure 9. *Cont.*

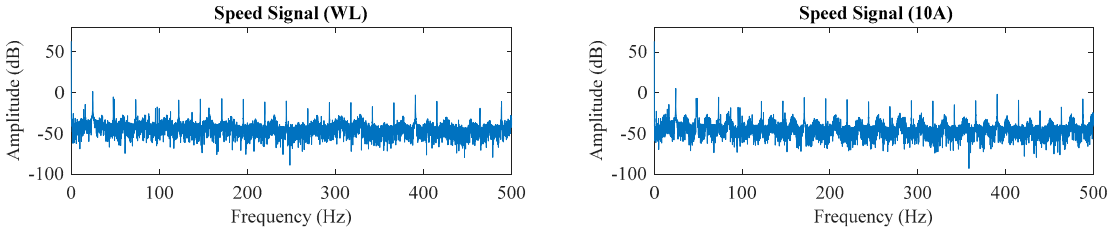


Figure 9. Log–Scale FFT Spectrums (in dB) of Current, Voltage, Torque, and Speed Signals Representing Frequency-Domain Characteristics of All Ten Bearing Fault Types in the Induction Machine Dataset.

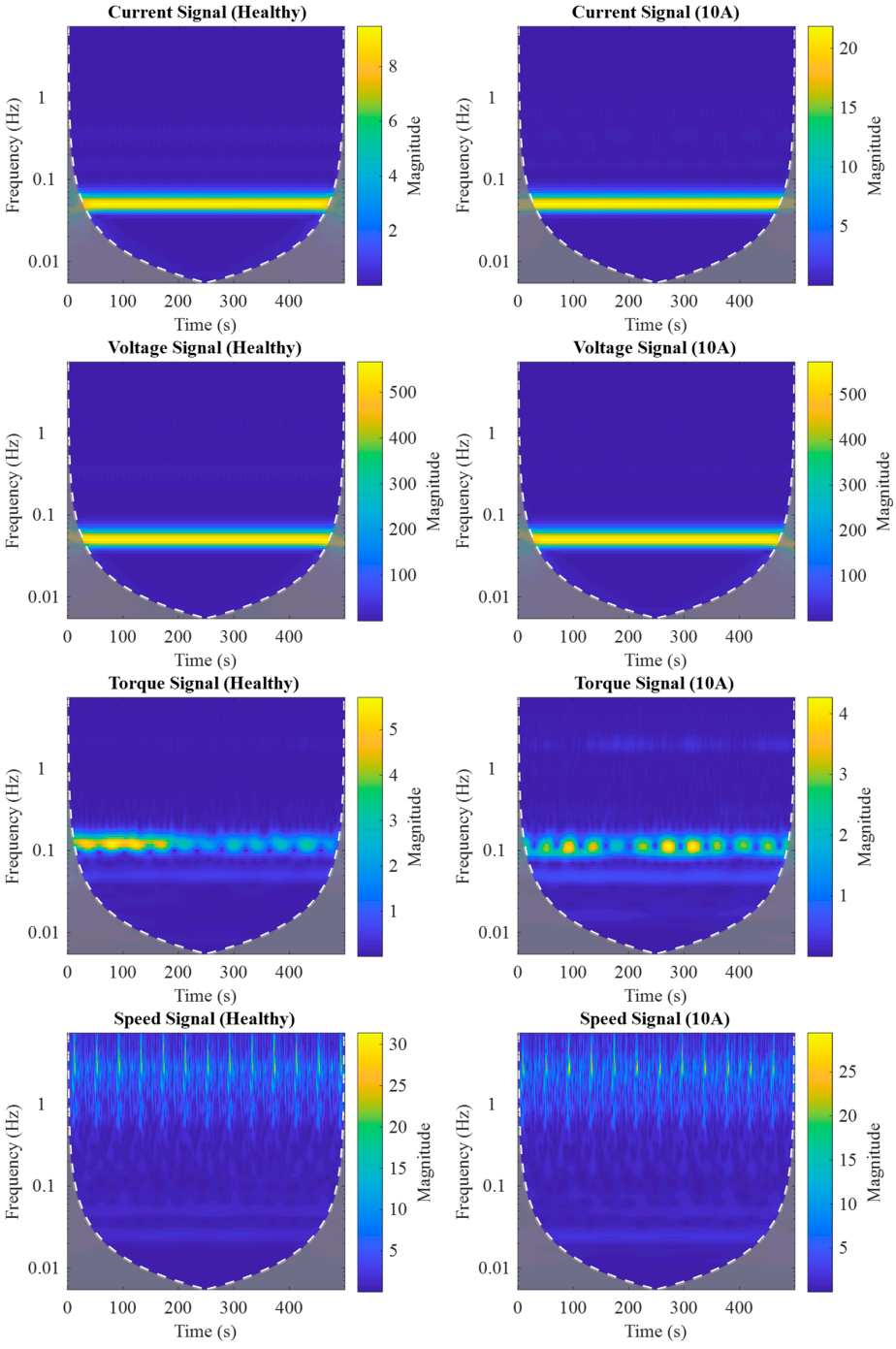


Figure 10. Time–Frequency CWT of Current, voltage, torque, and speed Signals Illustrating the Temporal Evolution of Frequency Components across Bearing Fault Types in the Induction Machine Dataset.

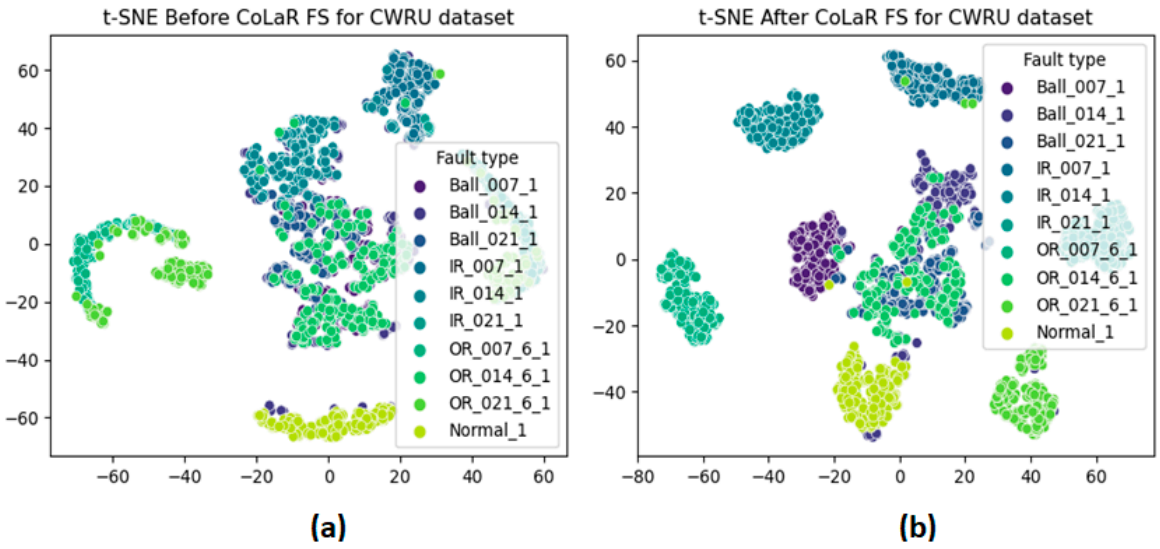


Figure 11. Feature distribution visualizations employing t-SNE for the CWRU bearing dataset (**a**) before applying the CoLaR-FS approach (**b**) after applying the CoLaR-FS approach.

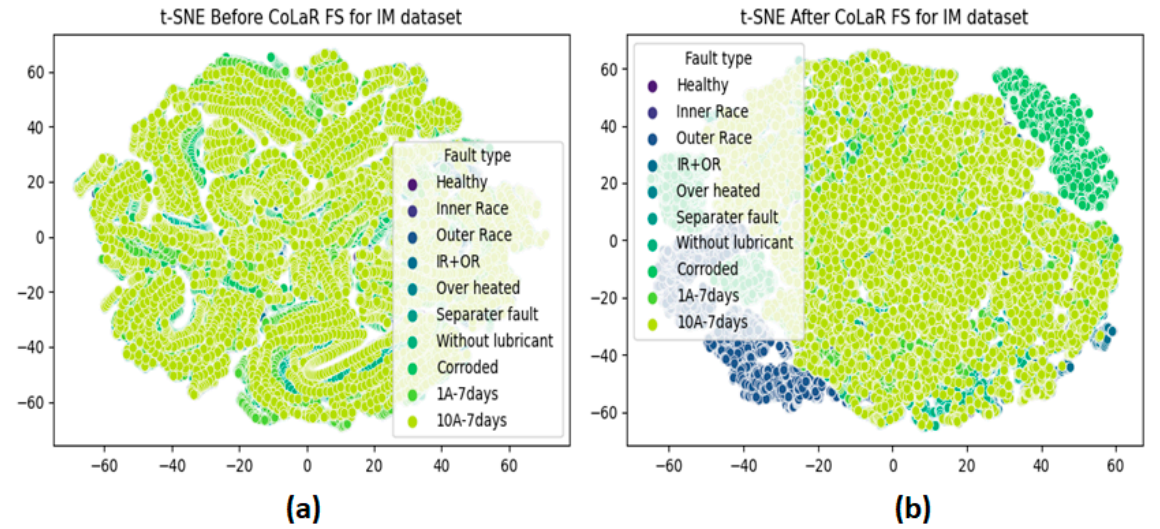


Figure 12. Feature distribution visualizations employing t-SNE for the IM bearing dataset (**a**) before applying the CoLaR-FS approach (**b**) after applying the CoLaR-FS approach.

The performance of the proposed fault detection and classification framework was evaluated using the CWRU dataset, both with and without the application of the CoLaR-FS feature selection algorithm, as shown in Figures 13 and 14. Confusion matrices and receiver operating characteristic (ROC) curves were utilized to assess the classification performance and the model's capability to discriminate among different fault types. Before applying CoLaR-FS, the model achieved a classification accuracy of 87.01%, as indicated by the corresponding confusion matrix and ROC curves. However, after employing CoLaR-FS to identify the most relevant features and eliminate redundant dimensions, the classification accuracy improved significantly to 98.35%. The enhanced ROC curves following feature selection confirmed this improvement, demonstrating better class separability and larger area-under-curve (AUC) values. These findings clearly indicate that integrating CoLaR-FS with the hybrid LSTM–XGBoost model substantially enhances fault detection and classification performance. The marked improvement underscores the importance of feature selection in achieving higher diagnostic accuracy and reliability when multiple fault symptoms coexist in rotating electrical machines [72].

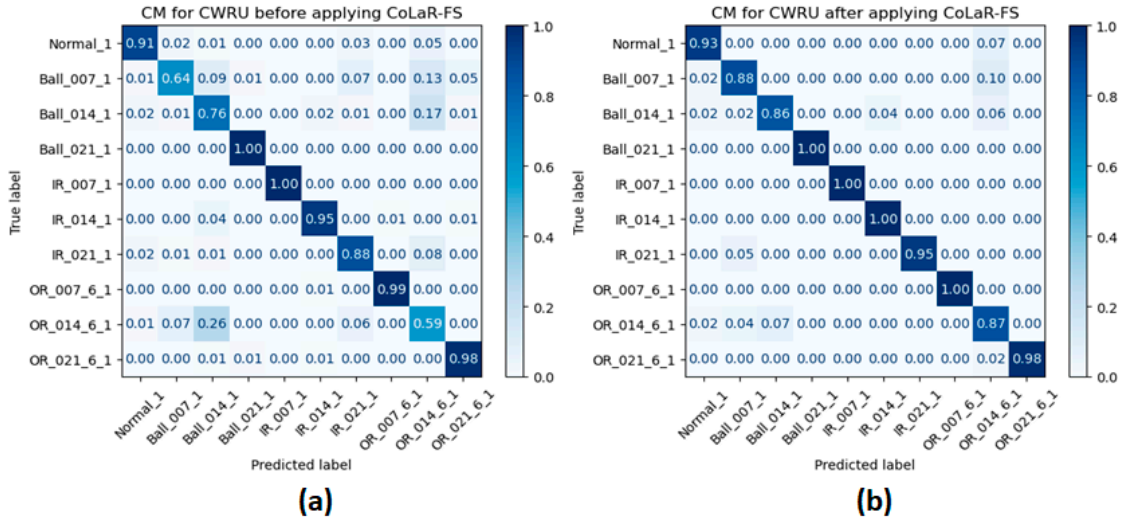


Figure 13. Confusion Matrix Showing RNN-XGBoost Classification Performance on the IM Dataset: (**a**) Before and (**b**) After Employing CoLaR-FS.

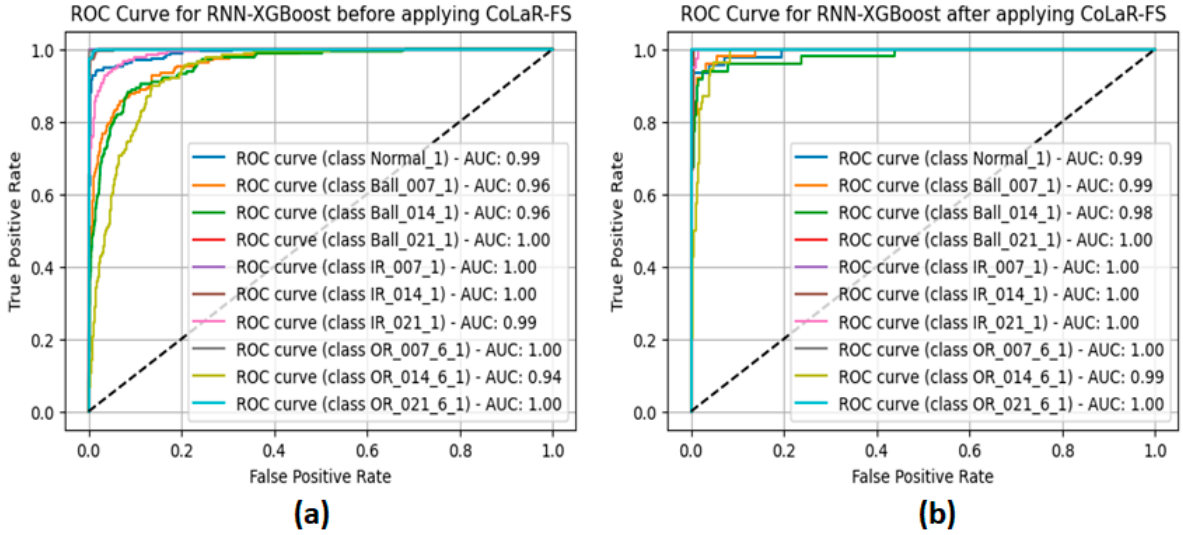


Figure 14. ROC Curves Illustrating RNN-XGBoost Classification Performance on the CWRU Dataset: (**a**) Before and (**b**) After Employing CoLaR-FS.

At the first baseline before the application of the CoLaR-FS, the proposed model yielded a classification accuracy of 79.98%. However, when using CoLaR-FS, which employs feature selection where only the most effective features with the least dimensionality are used, the accuracy shot up to 98.57%. The confusion matrices and the ROC curves vividly showed this improvement. In the case where feature selection was applied, the model was unable to differentiate between some of the faults; however, CoLaR-FS improved the separation of the ROC curves, improving the AUC; hence, providing a clearer classification of the different faults. This improvement shows that the CoLaR-FS method is useful and proves that applying it to the hybrid LSTM-XGBoost model will create a better and more efficient IM dataset fault detection and classification system. Figures 15 and 16 show the outcomes of the confusion matrix and ROC curves for the IM dataset with and without applying the proposed methodology, respectively.

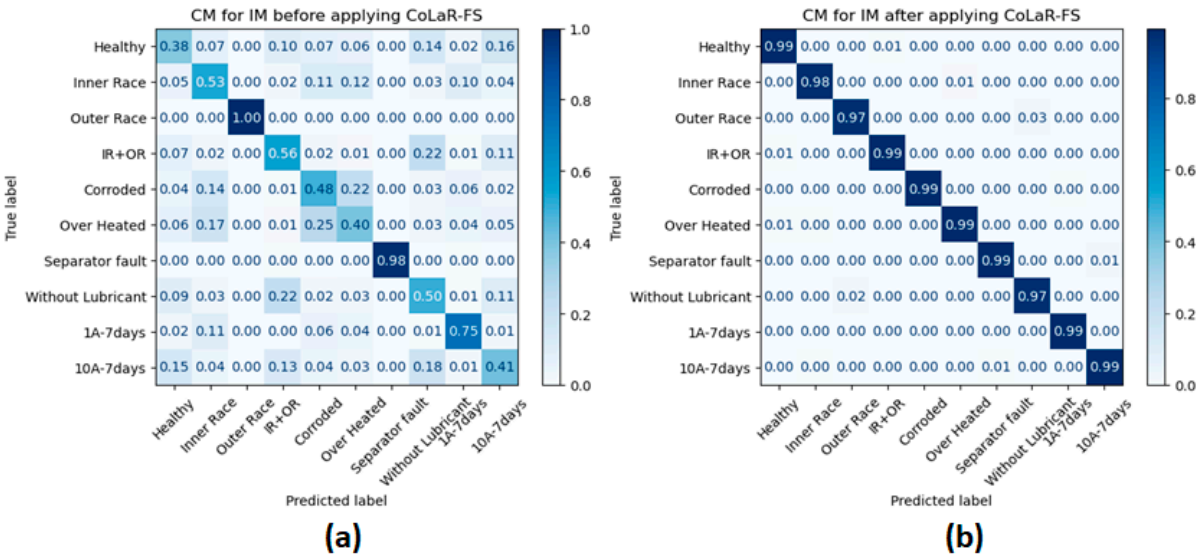


Figure 15. Confusion Matrix Showing RNN-XGBoost Classification Performance on the IM Dataset: (**a**) Before and (**b**) After Employing CoLaR-FS.

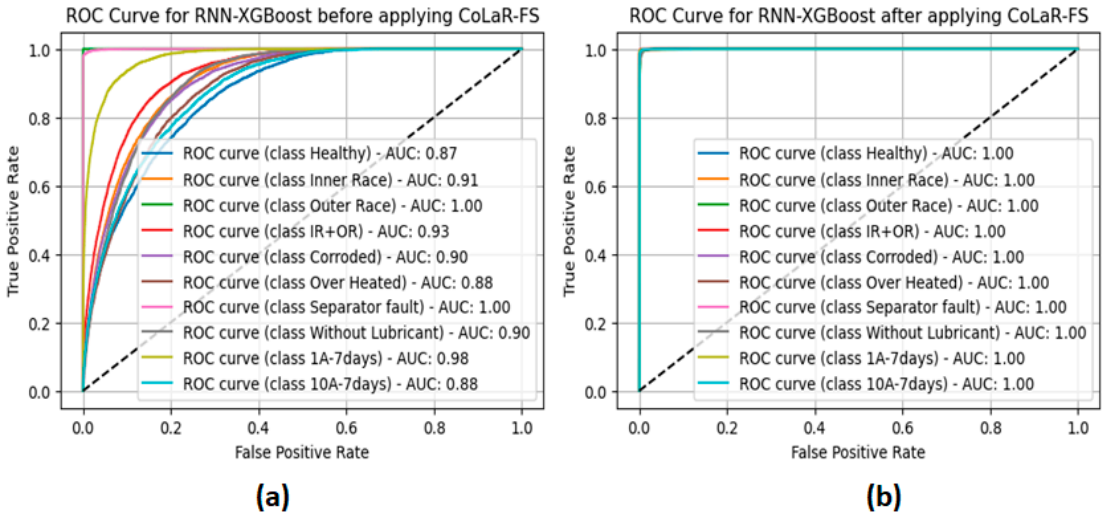


Figure 16. ROC Curves Illustrating RNN-XGBoost Classification Performance on the IM Dataset: (**a**) Before and (**b**) After Employing CoLaR-FS.

Tables 12 and 13 present the evaluation of the proposed RNN-XGBoost algorithm on the CWRU dataset, both pre- and post-CoLaR-FS feature selection methods. When CoLaR-FS was not included, the performance of the model was different for various types of faults: accuracies were from 61 to 99%. Precision, recall, and the F1 factor also showed some variance: there is no easy way to classify certain types of faults. When applied, the CoLaR-FS led to enhanced performance of the model in comparison with before. The optimal classification accuracy grew in all the fault kinds, reaching nearly ideal values in many cases. Also, precision, recall, and F1 scores were significantly higher, which proves the efficiency of the CoLaR-FS to improve the classification effectiveness of the classifier in terms of separating the different kinds of faults. Rating from the fact that the application of CoLaR-FS aided in the reduction of the input data into meaningful features, the general classification was enhanced due to concentration on the most important features.

Table 12. Performance Metrics of the proposed algorithm (RNN-XGBoost) on the CWRU Dataset without Applying CoLaR-FS.

Sr.No	Labels	Accuracy	Precision	Recall	F1 Score	Support
1	Normal_1	0.87	0.94	0.91	0.92	192
2	Ball_007	0.88	0.87	0.64	0.74	194
3	Ball_014	0.87	0.64	0.76	0.70	178
4	Ball_021	0.86	0.98	1.00	0.99	183
5	IR_007	0.99	1.00	1.00	1.00	172
6	IR_014	0.97	0.97	0.95	0.96	182
7	IR_021	0.96	0.84	0.88	0.86	180
8	OR_007	0.98	0.99	0.99	0.99	193
9	OR_014	0.61	0.57	0.59	0.58	178
10	OR_021	0.94	0.94	0.98	0.96	188

Table 13. Performance Metrics of the proposed algorithm (RNN-XGBoost) on the CWRU Dataset by Applying CoLaR-FS.

Sr.No	Labels	Accuracy	Precision	Recall	F1 Score	Support
1	Normal_1	0.95	0.93	0.93	0.93	192
2	Ball_007	0.94	0.94	0.88	0.89	194
3	Ball_014	0.95	0.95	0.86	0.88	178
4	Ball_021	1.00	1.00	1.00	1.00	183
5	IR_007	1.00	1.00	1.00	1.00	172
6	IR_014	0.97	0.97	1.00	0.97	182
7	IR_021	1.00	1.00	0.95	0.97	180
8	OR_007	1.00	1.00	1.00	1.00	193
9	OR_014	0.98	0.98	0.87	0.83	178
10	OR_021	1.00	1.00	0.98	0.99	188

Tables 14 and 15 display the efficiencies of the proposed RNN-XGBoost algorithm in the case of the IM dataset, while employing and without employing the CoLaR-FS feature selection technique. The results also depict that without the use of CoLaR-FS, the proposed algorithm performs differently over various kinds of faults, and the accuracy falls form 61% to 99%. Such faults as "Healthy" and "OH" yielded lower accuracy, precision, and F1-measure, which pointed to classification difficulties. However, comparatively high accuracy was observed for the faults like "OR" and "SF", as evident from Tables 3 and 4, where the precision and recall were not fully balanced, maybe due to the "OR" fault. By employing CoLaR-FS, the effectiveness of the model changed for the better across each of the fault categories. The precision, recall, accuracy, and F1 Measure were significantly high in most of the fault types as well, and accuracy and precision for many of them were nearly perfect. This suggests that the application of CoLaR-FS feature selection improved the capacity of the classifier in separating between faults, from the significant features only. For example, for most of the fault types, the accuracy improved to as high as 99%, and other measures such as precision and recall were also observed to improve. In general, the incorporation of CoLaR-FS enabled the model to obtain an improved base in generating more dependable and exact classifications of all faults contained in the database.

Table 14. Performance Metrics of the proposed algorithm (RNN-XGBoost) on IM Dataset without Applying CoLaR-FS.

Sr.No	Labels	Accuracy	Precision	Recall	F1 Score	Support
1	Healthy	0.61	0.59	0.61	0.62	4796
2	IR	0.77	0.76	0.72	0.74	4883
3	OR	0.99	0.98	0.99	0.68	4821
4	IR + OR	0.70	0.68	0.78	0.69	4773
5	Cd	0.79	0.74	078	1.00	4738
6	OH	0.69	0.67	0.75	0.67	4784
7	SF	0.98	0.99	0.98	0.99	4815
8	WL	0.71	0.69	0.69	0.71	4781
9	1A-7days	0.77	0.76	0.75	0.76	4841
10	10A-7days	0.69	0.64	0.70	0.73	4808

Table 15. Performance Metrics of the proposed algorithm (RNN-XGBoost) on IM Dataset by Applying CoLaR-FS.

Sr.No	Labels	Accuracy	Precision	Recall	F1 Score	Support
1	Healthy	0.99	0.99	0.99	0.99	4796
2	IR	0.99	0.99	0.98	0.99	4883
3	OR	0.98	0.97	0.97	0.97	4821
4	IR + OR	0.99	0.99	0.99	0.99	4773
5	Cd	0.99	0.99	0.99	0.99	4738
6	OH	0.99	0.99	0.99	0.99	4784
7	SF	0.99	0.99	0.99	0.99	4815
8	WL	0.98	0.97	0.97	097	4781
9	1A-7days	0.99	0.99	0.99	0.99	4841
10	10A-7days	0.99	0.99	0.99	0.99	4808

5.2. Computational Complexity and Lightweight Analysis

To validate the claimed lightweight nature of the proposed CoLaR-FS–RNN-XBoostNet framework, a quantitative analysis of model complexity and computational efficiency was conducted. In industrial condition monitoring applications, especially in real-time and edge-computing environments, metrics such as parameter count, computational cost, and inference latency are critical alongside diagnostic accuracy. The total number of trainable parameters in the proposed framework is primarily determined by the RNN-based feature extraction module, as the XGBoost classifier relies on tree-based decision structures rather than dense parameter matrices. Due to the prior application of CoLaR-FS, the dimensionality of the input feature space is significantly reduced before being fed into the learning architecture. From a computational perspective, the complexity analysis verifies that the proposed framework achieves a favorable balance between diagnostic accuracy and computational efficiency. The reduced parameter count and lower inference latency demonstrate that high diagnostic performance does not necessarily require deep or computationally intensive architectures. Instead, targeted feature optimization combined with lightweight temporal modeling can deliver superior results while maintaining efficiency, as shown in Table 16.

Table 16. Computational complexity comparison of bearing fault diagnosis models.

Model	Feature Dimension	Trainable Parameters	FLOPs (Approx.)	Inference Time per Sample (ms)	Diagnostic Accuracy (%)
CNN–LSTM	High	~1.2 M	High	8.7	97.95
XGBoost	Medium	-	Medium	4.1	95.61
RNN	Medium–High	~0.45 M	Medium	5.6	96.31
Proposed	Low (optimized)	~0.21 M	Low	2.8	98.57

Performance of several feature selection techniques with and without the proposed RNN-XGBoostNet model on two datasets, CWRU and IM, is compared in this section. Recursive Feature Elimination (RFE) is highly efficient for small data sets and relatively slow for large data sets. In data preprocessing, it assists in the removal of features that add minimal new information, but regarding features that are overlapping or correlated to a high degree, Lasso Regression is problematic. Among methods of feature reduction, the Principal Component Analysis (PCA) is highly effective in reducing features but bears the problem of interpretability. Exhaustive Feature Selection (EFS) provides a high non-linear relationship but is time-consuming for large data sets. MI measures feature relevance and also fails to deal with features that are correlated with each other appropriately. Compared with other methods, the proposed CoLaR-FS gives superior performance when applied to hybrid models such as RNN-XGBoostNet. When comparing the results, a higher accuracy of CoLaR-FS showed the efficiency of the method for bearing fault diagnosis in the IM dataset, as shown in Table 17.

Table 17. Comparison of Feature Selection Techniques on Bearing Fault Diagnosis Using RNN-XGBoostNet.

Feature Selection Method	Selected Features	Classifier Used	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Correlation Filtering	All (F1–F8)	RNN-XBoostNet	92.51	91.88	92.05	91.96
Lasso Regression	F1, F2, F4, F5, F6	RNN-XBoostNet	93.24	92.71	92.58	92.69
Recursive Feature Elimination (RFE)	F1, F3, F4, F6, F7	RNN-XBoostNet	94.11	93.88	93.54	93.61
Mutual Information (MI)	F1, F2, F3, F5, F6, F8	RNN-XBoostNet	94.58	94.04	93.81	93.97
Principal Component Analysis (PCA)	F1, F2, F4, F5, F6, F8	RNN-XBoostNet	93.89	93.22	93.05	93.15
CoLaR-FS (Proposed)	F1, F2, F3, F4, F6	RNN-XBoostNet	99.57	99.54	99.21	99.37

6. Conclusions

This work introduced RNN-XBoostNet, an advanced hybrid diagnostic framework supported by the CoLaR-FS, a feature-selection approach for reliable bearing fault detection in rotating electrical machines under varying circumstances. By integrating temporal feature extraction through RNNs to strengthen the classification strength of the extremely influential ensemble learning algorithm XGBoost, the proposed method effectively addressed challenges related to noisy, high-dimensional, and non-stationary data. Extensive evaluations on both the benchmark CWRU dataset and the experimentally collected IM

dataset demonstrated significant improvements in classification accuracy, underscoring the model's robustness, generalization ability, and practical applicability in real-world industrial monitoring scenarios. Building on these promising results, future developments may focus on deploying the framework in real-time embedded and edge-based platforms, integrating multi-sensor data fusion for richer diagnostics, and adopting transfer learning to reduce data requirements when adapting to new machines or fault types. Further enhancements through explainable AI (XAI) could also increase transparency for maintenance engineers, while extending the framework toward fault-severity estimation and predictive maintenance scheduling would strengthen its industrial relevance. Overall, the proposed methodology provides a solid foundation for next-generation intelligent condition-monitoring systems.

Author Contributions: Conceptualization, M.A.K., M.U.N. and B.A.; methodology, M.A.K.; software, M.A.K.; validation, B.A. and T.V.; formal analysis, T.V.; investigation, M.A.K. and B.A.; resources, M.A.K.; data curation, M.U.N.,B.A. and T.V.; writing—original draft preparation, M.A.K. and B.A.; writing—review and editing, B.A.; visualization, B.A.; supervision, T.V. and A.K.All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chen, F.; Zhou, X.; Xu, B.; Yang, Z.; Qu, Z. Instantaneous square current signal analysis for motors using vision transformer for the fault diagnosis of rolling bearings. *Appl. Sci.* **2023**, *13*, 9349. [CrossRef]
2. Li, Z.; Wang, Y.; Ma, J. Fault diagnosis of motor bearings based on a convolutional long short-term memory network of bayesian optimization. *IEEE Access* **2021**, *9*, 97546–97556. [CrossRef]
3. Toma, R.N.; Prosvirin, A.E.; Kim, J.M. Bearing fault diagnosis of induction motors using a genetic algorithm and machine learning classifiers. *Sensors* **2020**, *20*, 1884. [CrossRef]
4. Wang, X.; Meng, R.; Wang, G.; Liu, X.; Liu, X.; Lu, D. The research on fault diagnosis of rolling bearing based on current signal CNN-SVM. *Meas. Sci. Technol.* **2023**, *34*, 125021. [CrossRef]
5. Ye, X.; Li, G. An intelligent fault diagnosis method for rolling bearing using motor stator current signals. *Meas. Sci. Technol.* **2024**, *35*, 086131. [CrossRef]
6. Zhang, B.; Li, H.; Kong, W.; Fu, M.; Ma, J. Early-stage fault diagnosis of motor bearing based on kurtosis weighting and fusion of current–vibration signals. *Sensors* **2024**, *24*, 3373. [CrossRef]
7. Zhou, S.; Lin, L.; Chen, C.; Pan, W.; Lou, X. Application of convolutional neural network in motor bearing fault diagnosis. *Comput. Intell. Neurosci.* **2022**, *2022*, 9231305. [CrossRef]
8. Belmiloud, D.; Benkedjouh, T.; Lachi, M.; Laggoun, A.; Dron, J. Deep convolutional neural networks for bearings failure predictionand temperature correlation. *J. Vibroeng.* **2018**, *20*, 2878–2891. [CrossRef]
9. Che, C.; Wang, H.; Lin, R.; Ni, X. Deep meta-learning and variational autoencoder for coupling fault diagnosis of rolling bearing under variable working conditions. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2022**, *236*, 9900–9913. [CrossRef]
10. Gao, Z.; Xi, C.; Yang, Y. Fault diagnosis of wind turbine bearings based on VMD-Tsallis entropy and SVM. In Proceedings of the Eighth International Conference on Energy Materials and Electrical Engineering (ICEMEE 2022), Zhangjiajie, China, 22–24 April 2022; SPIE: Bellingham, WA, USA, 2023. [CrossRef]
11. Gu, Y.; Cao, J.; Song, X.; Yao, J. A denoising autoencoder-based bearing fault diagnosis system for time-domain vibration signals. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 9790053. [CrossRef]
12. Hamadache, M.; Jung, J.; Park, J.; Youn, B. A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: Shallow and deep learning. *JMST Adv.* **2019**, *1*, 125–151. [CrossRef]

13. Mushtaq, S.; Islam, M.; Sohaib, M. Deep learning aided data-driven fault diagnosis of rotatory machine: A comprehensive review. *Energies* **2021**, *14*, 5150. [CrossRef]
14. Pham, M.; Kim, J.; Kim, C. Intelligent fault diagnosis method using acoustic emission signals for bearings under complex working conditions. *Appl. Sci.* **2020**, *10*, 7068. [CrossRef]
15. Saufi, M.; Ahmad, Z.; Leong, M.; Lim, M. Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. *IEEE Access* **2019**, *7*, 122644–122662. [CrossRef]
16. Wang, J.; Mo, Z.; Zhang, H.; Miao, Q. A deep learning method for bearing fault diagnosis based on time-frequency image. *IEEE Access* **2019**, *7*, 42373–42383. [CrossRef]
17. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
18. Ge, Z.; Cao, Y.; Luo, J.; Yang, J. Fault detection of spindle rolling bearings in CNC machine tools based on odorless Kalman filtering algorithm. In *Intelligent Computing Technology and Automation*; IOS Press: Amsterdam, The Netherlands, 2024. [CrossRef]
19. Li, C.; Liu, Y.; Liao, Y. An improved parameter-adaptive variational mode decomposition method and its application in fault diagnosis of rolling bearings. *Shock. Vib.* **2021**, *2021*, 2968488. [CrossRef]
20. Luo, Y.; Lu, W.; Kang, S.; Tian, X.; Kang, X.; Sun, F. Enhanced feature extraction network based on acoustic signal feature learning for bearing fault diagnosis. *Sensors* **2023**, *23*, 8703. [CrossRef]
21. Wu, S.; Zhou, J.; Liu, T. Compound fault feature extraction of rolling bearing acoustic signals based on AVMD-IMVO-MCKD. *Sensors* **2022**, *22*, 6769. [CrossRef]
22. Xu, X.; Li, Z.; Zhao, Y.; Qian, W. Fault diagnosis of gas turbine generator bearings using enhanced valuable sample strategy and convolutional neural network. *Meas. Sci. Technol.* **2023**, *35*, 015021. [CrossRef]
23. Zheng, J.; Yuan, Y.; Zou, L.; Deng, W.; Chen, G.; Zhao, H. Study on a novel fault diagnosis method based on VMD and BLM. *Symmetry* **2019**, *11*, 747. [CrossRef]
24. Zhou, X.; Li, A.; Han, G. An intelligent multi-local model bearing fault diagnosis method using small sample fusion. *Sensors* **2023**, *23*, 7567. [CrossRef]
25. Zhang, Q.; Deng, L. An intelligent fault diagnosis method of rolling bearings based on short-time Fourier transform and convolutional neural network. *J. Fail. Anal. Prev.* **2023**, *23*, 795–811. [CrossRef]
26. Nasser Mohamed, Y.; Seker, S.; Akinci, T.C. Signal processing application based on a hybrid wavelet transform to fault detection and identification in power system. *Information* **2023**, *14*, 540. [CrossRef]
27. Liu, S.; Sun, Y.; He, L.; Kang, Y. Weak signal processing methods based on improved HHT and filtering techniques for steel wire rope. *Appl. Sci.* **2022**, *12*, 6969. [CrossRef]
28. Amanuel, T.; Ghirmay, A.; Ghebremeskel, H.; Ghebrehiwet, R.; Bahlibi, W. Comparative analysis of signal processing techniques for fault detection in three phase induction motor. *J. Electron.* **2021**, *3*, 61–76.
29. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [CrossRef]
30. Soomro, A.A.; Muhammad, M.B.; Mokhtar, A.A.; Saad, M.H.M.; Lashari, N.; Hussain, M.; Sarwar, U.; Palli, A.S. Insights into modern machine learning approaches for bearing fault classification: A systematic literature review. *Results Eng.* **2024**, *23*, 102700. [CrossRef]
31. Gougam, F.; Afia, A.; Soualhi, A.; Touzout, W.; Rahmoune, C.; Benazzouz, D. Bearing faults classification using a new approach of signal processing combined with machine learning algorithms. *J. Braz. Soc. Mech. Sci. Eng.* **2024**, *46*, 65. [CrossRef]
32. Vishwendra, M.A.; Salunkhe, P.S.; Patil, S.V.; Shinde, S.A.; Shinde, P.V.; Desavale, R.G.; Jadhav, P.M.; Dharwadkar, N.V. A novel method to classify rolling element bearing faults using K-nearest neighbor machine learning algorithm. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* **2022**, *8*, 031202. [CrossRef]
33. Wang, H.; Zheng, J.; Xiang, J. Online bearing fault diagnosis using numerical simulation models and machine learning classifications. *Reliab. Eng. Syst. Saf.* **2023**, *234*, 109142. [CrossRef]
34. Sinitsin, V.; Ibryaeva, O.; Sakovskaya, V.; Eremeeva, V. Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model. *Mech. Syst. Signal Process.* **2022**, *180*, 109454. [CrossRef]
35. Wang, Y.; Li, D.; Li, L.; Sun, R.; Wang, S. A novel deep learning framework for rolling bearing fault diagnosis enhancement using VAE-augmented CNN model. *Heliyon* **2024**, *10*, e35407. [CrossRef] [PubMed]
36. Joung, B.G.; Nath, C.; Li, Z.; Sutherland, J.W. Bearing anomaly detection in an air compressor using an LSTM and RNN-based machine learning model. *Int. J. Adv. Manuf. Technol.* **2024**, *134*, 3519–3530. [CrossRef]
37. Toma, R.N.; Piltan, F.; Kim, J.M. A deep autoencoder-based convolution neural network framework for bearing fault classification in induction motors. *Sensors* **2021**, *21*, 8453. [CrossRef] [PubMed]
38. Garavaso, L.O. Enhancing LightGBM Fault Prediction in Rotating Machines Using Knowledge-Based Resampling and Bayesian Optimization. Master's Thesis, Universida de Estadual De Campinas, Campinas, Brazil, 2024.

39. Lee, C.Y.; Maceren, E.D.C. Induction motor bearing fault classification using deep neural network with particle swarm optimization-extreme gradient boosting. *IET Electr. Power Appl.* **2024**, *18*, 297–311. [CrossRef]
40. Lu, S.; Lu, J.; An, K.; Wang, X.; He, Q. Edge computing on IoT for machine signal processing and fault diagnosis: A review. *IEEE Internet Things J.* **2023**, *10*, 11093–11116. [CrossRef]
41. Wang, M.; Yu, J.; Leng, H.; Du, X.; Liu, Y. Bearing fault detection by using graph autoencoder and ensemble learning. *Sci. Rep.* **2024**, *14*, 5206. [CrossRef]
42. Chen, H.; Luo, H.; Huang, B.; Jiang, B.; Kaynak, O. Transfer learning-motivated intelligent fault diagnosis designs: A survey, insights, and perspectives. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 2969–2983. [CrossRef]
43. Sanakkayala, D.C.; Varadarajan, V.; Kumar, N.; Karan; Soni, G.; Kamat, P.; Kumar, S.; Patil, S.; Kotecha, K. Explainable AI for bearing fault prognosis using deep learning techniques. *Micromachines* **2022**, *13*, 1471. [CrossRef]
44. Malhotra, A.; Minhas, A.S.; Singh, S.; Zuo, M.J.; Kumar, R.; Kankar, P.K. Bearing fault diagnosis based on flexible analytical wavelet transform and fuzzy entropy approach. *Mater. Today Proc.* **2021**, *43*, 629–635. [CrossRef]
45. Tian, H.; Fan, H.; Feng, M.; Cao, R.; Li, D. Fault diagnosis of rolling bearing based on HPSO algorithm optimized CNN-LSTM neural network. *Sensors* **2023**, *23*, 6508. [CrossRef]
46. Xia, X.; Sun, H.; Wang, A. Fault Diagnosis of Planetary Gearboxes Based on LSTM Improved via Feature Extraction Using VMD, Fusion Entropy, and Random Forest. *Entropy* **2025**, *27*, 956. [CrossRef]
47. Ding, Y.; Jia, M.; Miao, Q.; Cao, Y. A novel time–frequency Transformer based on self–attention mechanism and its application in fault diagnosis of rolling bearings. *Mech. Syst. Signal Process.* **2022**, *168*, 108616. [CrossRef]
48. Jain, P.H.; Bhosle, S.P. Analysis of vibration signals caused by ball bearing defects using time-domain statistical indicators. *Int. J. Adv. Technol. Eng. Explor.* **2022**, *9*, 700. [CrossRef]
49. Liang, P.; Wang, W.; Yuan, X.; Liu, S.; Zhang, L.; Cheng, Y. Intelligent fault diagnosis of rolling bearing based on wavelet transform and improved ResNet under noisy labels and environment. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105269. [CrossRef]
50. Jin, Z.; He, D.; Wei, Z. Intelligent fault diagnosis of train axle box bearing based on parameter optimization VMD and improved DBN. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104713. [CrossRef]
51. Huang, Y.J.; Liao, A.H.; Hu, D.Y.; Shi, W.; Zheng, S.B. Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis. *Measurement* **2022**, *203*, 111935. [CrossRef]
52. Kumar, D.; Ujjan, S.M.; Dev, K.; Khowaja, S.A.; Bhatti, N.A.; Hussain, T. Towards soft real-time fault diagnosis for edge devices in industrial IoT using deep domain adaptation training strategy. *J. Parallel Distrib. Comput.* **2022**, *160*, 90–99. [CrossRef]
53. Li, Z.; Zhang, K.; Zheng, Q.; Ding, G.; Hao, W.; Zhang, H.; Zhang, W. Unsupervised fault detection with multi-source anomaly sensitivity enhancing convolutional autoencoder for high-speed train bogie bearings. *Expert Syst. Appl.* **2025**, *281*, 127570. [CrossRef]
54. Wang, J.; Wang, D.; Wang, S.; Li, W.; Song, K. Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network. *IEEE Access* **2021**, *9*, 23717–23725. [CrossRef]
55. Zhong, H.; Lv, Y.; Yuan, R.; Yang, D. Bearing fault diagnosis using transfer learning and self-attention ensemble lightweight convolutional neural network. *Neurocomputing* **2022**, *501*, 765–777. [CrossRef]
56. Brito, L.C.; Susto, G.A.; Brito, J.N.; Duarte, M.A.V. Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction. *Informatics* **2021**, *8*, 85. [CrossRef]
57. Rohan, A.I.; Ridita, T.A.; Anonto, H.Z.; Hossain, M.I.; Shufian, A.; Mahin, M.S.R.; Islam, M.A. Intelligent Fault Diagnosis in Rolling Element Bearings: Combining Envelope Spectrum and Spectral Kurtosis for Enhanced Detection. *Results Eng.* **2025**, *27*, 106899. [CrossRef]
58. Feng, Z.; Zhang, Z. A small sample rolling bearing fault diagnosis based on PSD-VME and DS evidence theory enhancedmRVM. *Comput. Electr. Eng.* **2024**, *118*, 109458. [CrossRef]
59. Deng, X.; Li, M.; Wang, L.; Wan, Q. Rfcbf: Enhance the performance and stability of fast correlation-based filter. *Int. J. Comput. Intell. Appl.* **2022**, *21*, 2250009. [CrossRef]
60. Li, Y.; Lu, F.; Yin, Y. Applying logistic LASSO regression for the diagnosis of atypical Crohn's disease. *Sci. Rep.* **2022**, *12*, 11340. [CrossRef] [PubMed]
61. Kornyo, O.; Asante, M.; Opoku, R.; Owusu-Agyemang, K.; Partey, B.T.; Baah, E.K.; Boadu, N. Botnet attacks classification in AMI networks with recursive feature elimination (RFE) and machine learning algorithms. *Comput. Secur.* **2023**, *135*, 103456. [CrossRef]
62. Ramezan, C.A. Transferability of recursive feature elimination (RFE)-derived feature sets for support vector machine land cover classification. *Remote Sens.* **2022**, *14*, 6218. [CrossRef]
63. Kanwal, I.; Wahid, F.; Ali, S.; Rehman, A.U.; Alkhayyat, A.; Al-Radaei, A. Sentiment analysis using hybrid model of stacked Auto-Encoder-Based feature extraction and long short term Memory-Based classification approach. *IEEE Access* **2023**, *11*, 124181–124197. [CrossRef]
64. Li, W.; Fu, H.; Han, Z.; Zhang, X.; Jin, H. Intelligent tool wear prediction based on Informer encoder and stacked bidirectional gated recurrent unit. *Robot. Comput.-Integr. Manuf.* **2022**, *77*, 102368. [CrossRef]

65. Kavzoglu, T.; Teke, A. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arab. J. Sci. Eng.* **2022**, *47*, 7367–7385. [CrossRef]
66. Ali, Y.A.; Awwad, E.M.; Al-Razgan, M.; Maarouf, A. Hyperparameter search for machine learning algorithms for optimizing the computational complexity. *Processes* **2023**, *11*, 349. [CrossRef]
67. Zhang, X.; Zhao, B.; Lin, Y. Machine learning based bearing fault diagnosis using the case Western Reserve university data: A review. *IEEE Access* **2021**, *9*, 155598–155608. [CrossRef]
68. Raj, K.K.; Kumar, S.; Kumar, R.R.; Andriollo, M. Enhanced fault detection in bearings using machine learning and raw accelerometer data: A case study using the Case Western Reserve University dataset. *Information* **2024**, *15*, 259. [CrossRef]
69. Bhat, S.S.; Ansari, G.A. Machine learning algorithm-based early prediction of diabetes: A new feature selection using correlation matrix with heat map. In *Combating Women's Health Issues with Machine Learning*; CRC Press: Boca Raton, FL, USA, 2023; pp. 184–198.
70. Ribeiro Junior, R.F.; Areias, I.A.D.S.; Gomes, G.F. Fault detection and diagnosis using vibration signal analysis in frequency domain for electric motors considering different real fault types. *Sens. Rev.* **2021**, *41*, 311–319. [CrossRef]
71. Liu, H.; Yang, J.; Ye, M.; James, S.C.; Tang, Z.; Dong, J.; Xing, T. Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. *J. Hydrol.* **2021**, *597*, 126146. [CrossRef]
72. Chicco, D.; Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min.* **2023**, *16*, 4. [CrossRef]
73. Naidu, G.; Zuva, T.; Sibanda, E.M. A review of evaluation metrics in machine learning algorithms. In *Computer Science On-Line Conference*; Springer International Publishing: Cham, Switzerland, 2023; pp. 15–25.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.