PAMUKKALE ÜNİVERSİTESİ
Mühendislik Bilimleri Dergisi

PAMUKKALE UNIVERSITY
Journal of Engineering Sciences

Araştırma makalesi | Research paper

Cilt | Vol: XX • Sayı | No: X • ss | pp: XXXX-XXXX • Yıl | Year: 20XX

Integrating multiscale entropy feature extraction with lightweight classifiers for robust bearing fault diagnosis in industrial systems

Derya Deliktaş [1], Gurkan Ozturk [2], Abdurrahman Ünsal [3], Ozden Ustun [4]

ARTICLE INFO

Dates:
Received: 10.05.2026
Accepted: 02.09.2026

Doi:
10.65206/pajes.1948523

Corresponding author:
Derya Deliktaş
(derya.deliktas@dpu.edu.tr; derya.deliktas@outlook.com)

Author addresses:
[1] Kütahya Dumlupınar University, Faculty of Engineering, Department of Industrial Engineering, Kütahya, Türkiye (derya.deliktas@dpu.edu.tr; derya.deliktas@outlook.com)

[2] Eskişehir Technical University, Faculty of Engineering, Department of Industrial Engineering, Eskişehir, Türkiye (gurkan.o@eskisehir.edu.tr)

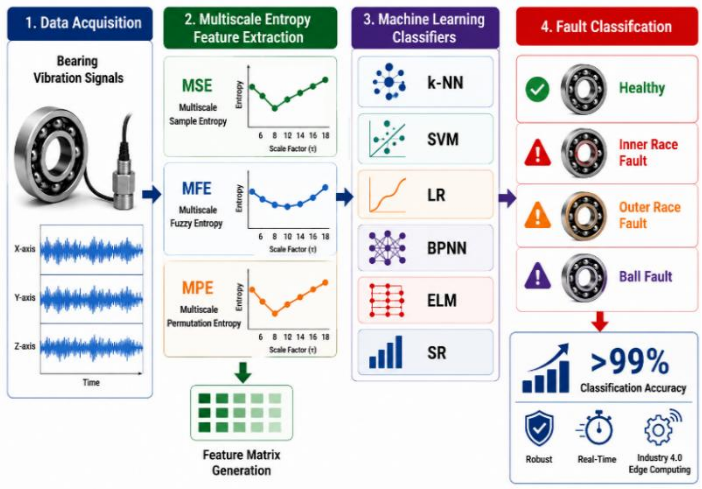
[3] Kütahya Dumlupınar University, Faculty of Engineering, Department of Electrical Electronics Engineering, Kütahya, Türkiye (unsal@dpu.edu.tr)

[4] Aksaray University, Faculty of Engineering, Department of Industrial Engineering, Aksaray, Türkiye (ozden.ustun@aksaray.edu.tr)

ABSTRACT

Context—Rolling-element bearings are among the most critical components in industrial machinery, and their failures may result in unexpected downtime, economic losses, and safety risks. As Industry 4.0 applications continue to expand, there is an increasing demand for reliable, computationally efficient, and real-time fault diagnosis systems capable of operating under varying industrial conditions.

Objective—This study aims to develop a robust and lightweight bearing fault diagnosis framework by integrating multiscale entropy-based feature extraction methods with conventional machine learning classifiers. The objective is to systematically evaluate the effectiveness, robustness, and parameter sensitivity of different entropy measures for accurate bearing fault classification under diverse operating conditions while maintaining suitability for edge-computing and real-time industrial monitoring applications.



Method—The proposed framework combines three multiscale entropy methods—Multiscale Sample Entropy, Multiscale Fuzzy Entropy, and Multiscale Permutation Entropy—with six lightweight classification algorithms: *k*-Nearest Neighbors, Support Vector Machine, Logistic Regression, Backpropagation Neural Network, Extreme Learning Machine, and Softmax Regression. Experimental evaluations were conducted using two benchmark datasets: the Case Western Reserve University (CWRU) drive-end bearing dataset and the CISAR three-axis industrial vibration dataset collected under realistic conditions. Hyperparameter selection was performed through stratified 10-fold cross-validation to ensure fair model comparison and prevent overfitting. In addition, sensitivity analysis was conducted by varying the entropy scale factor between 6 and 18 to investigate its effect on classification performance and feature stability.

Results—Experimental findings demonstrate consistently high classification performance across different fault types, fault diameters, and operating conditions. Among the investigated feature extraction methods, MFE showed comparatively high classification accuracy and stable performance across the experimental settings, frequently exceeding 99% classification accuracy across multiple classifier combinations. Several classifier-feature pairs achieved perfect classification performance on the CWRU dataset. The sensitivity analysis revealed that the optimal scale-factor range lies between $\tau = 12$ and 14. Compared with MSE and MPE, MFE showed descriptively lower sensitivity to scale-factor variation, indicating stronger robustness against operational variability and signal noise.

Conclusion—The proposed framework provides an accurate and lightweight approach for bearing fault diagnosis, with potential applicability to resource-constrained monitoring environments. The integration of multiscale entropy features with lightweight classifiers enables reliable predictive maintenance under varying industrial conditions. The results highlight the empirically higher performance of MFE-based feature extraction and provide practical guidance for developing robust Industry 4.0 fault monitoring systems.

Key Words—Bearing fault diagnosis, Extreme learning machine, Feature extraction, Multiscale entropy, Scale–factor sensitivity, Vibration analysis.

I. INTRODUCTION

In the digitalization era, industrial machines and cyber-physical systems play an essential role in industrial and service systems. Consequently, fault diagnosis and maintenance have received considerable attention. While corrective maintenance after failure can lead to high operating costs, proactive maintenance aims to prevent unplanned downtime by predicting failures before they occur. Automatic fault detection and identification are therefore essential for reliable machine operation [1]. Bearings are critical machine components exposed to changing loads, speeds, and operating conditions. Approximately 50% of common failures in electric motors are related to bearing problems [2], which can cause mechanical damage, production downtime, and safety risks. Therefore, effective bearing fault diagnosis is essential for improving the reliability and safety of mechanical equipment [3]. Common bearing defects include outer-race, inner-race, ball or roller, and cage defects, as well as multiple-component defects, looseness, misalignment, and inadequate lubrication.

Bearing fault diagnosis has attracted increasing research attention due to the growing demand for safety and reliability in industrial systems. Existing approaches can generally be classified as model-based and data-driven methods [4]–[7]. Model-based methods depend on analytical models and may be sensitive to changes in operating conditions [6]. Data-driven techniques, in contrast, use historical data without requiring an explicit mathematical model, making them suitable for changing environments [8]–[11].

Machine learning methods, including artificial neural networks, support vector machines, and principal component analysis, have been widely applied to bearing fault diagnosis [12], [13]. A typical fault classification framework includes data collection, feature extraction, and fault identification [14]. However, separate feature extraction and classification stages may require manual feature engineering, increasing complexity and computational time [15]. Statistical approaches may also be sensitive to sensor placement and changing operating conditions [16], [17]. With the rapid development of artificial intelligence, deep learning methods have increasingly been applied to bearing fault diagnosis [18]–[21]. Early studies mainly employed vibration signals in deep learning frameworks [22], [23]. For example, Short-Time Fourier Transform and a Convolutional Neural Network have been combined and achieved approximately 97% accuracy [24], while a multi-signal fusion approach based on an improved deep convolutional generative adversarial network for imbalanced data has been proposed, achieving approximately 98% accuracy [25].

Current-signal (CS)-based diagnostic methods have also gained attention because CSs are easy to collect, inexpensive, and non-invasive [26], [27]. However, weak fault characteristics in CSs may limit diagnostic performance in practical applications [28]. A hybrid fault diagnostics and prognostics system has been proposed for motor-bearing health monitoring under multiple operating conditions, although it was limited to specific fault detection [29]. Several public datasets, including PRONOSTIA [30], CWRU [31], IMS [32], and Paderborn [33], are also widely used in data-driven bearing fault diagnosis.

It is important to clarify that the present study does not claim methodological novelty in the mathematical formulations of Multiscale Sample Entropy (MSE), Multiscale Fuzzy Entropy (MFE), or Multiscale Permutation Entropy (MPE). These methods are established entropy-based representations that have previously been introduced and applied in different signal analysis and fault diagnosis problems. The contribution of this study instead lies in providing a unified and controlled benchmarking framework for their comparative evaluation in bearing fault detection.

In contrast to studies that primarily investigate a single entropy representation or evaluate entropy-based features under different experimental configurations, the present study evaluates MSE, MFE, and MPE under the same signal segmentation, feature extraction, sampling, and cross-validation procedures. This controlled design enables the relative discriminative capability of the three entropy representations to be assessed without introducing performance differences arising from inconsistent experimental settings. Furthermore, the three entropy methods are systematically combined with six conventional and computationally lightweight classifiers, allowing the interaction between feature representation and classifier selection to be examined.

Another distinguishing aspect of the study is the evaluation of the entropy representations on two vibration datasets with different acquisition characteristics. The CWRU drive-end bearing dataset and the CISAR three-axis vibration dataset provide complementary experimental settings for examining the consistency of entropy-based representations across different signal characteristics. In addition, a scale-factor sensitivity analysis is conducted by varying the scale factor from 6 to 18 for all three entropy methods. This analysis provides further insight into the robustness of each entropy representation with respect to parameter variation and identifies an empirically favorable range associated with relatively stable classification performance.

Therefore, the novelty of the present study should be understood as a systematic comparative and benchmarking contribution rather than the introduction of a new entropy formulation. By jointly considering three multiscale entropy representations, six classifiers, two vibration datasets, and scale-factor sensitivity under a common experimental protocol, the study provides comparative evidence regarding the accuracy, stability, and practical robustness of MSE, MFE, and MPE for bearing fault detection. The current study,

- Established a unified benchmark for comparing three established multiscale entropy representations (MSE, MFE, and MPE) under the same preprocessing and validation conditions.
- Performed scale-factor sensitivity analysis (τ = 6-18), identifying the empirically favorable range (τ = 12-14) for stable classification performance.
- Evaluated the three entropy representations in combination with six lightweight classifiers to systematically compare their classification effectiveness.
- Validated the comparative performance of the entropy-based representations using both the CWRU drive-end bearing dataset and the CISAR three-axis vibration dataset.
- Provided practical insights into entropy selection, parameter sensitivity, and lightweight classifier combinations for reliable bearing fault detection.

The structure of this study is delineated as follows: Section II describes the implementation of bearing faults and the data collection procedures for the two benchmark datasets. Section III details the proposed fault classification framework, including the entropy-based feature extraction methods and the six classification algorithms. Section IV presents comparative benchmarking results together with the scale-factor sensitivity analysis. Section V offers a comprehensive discussion of the findings and concluding remarks.

II. IMPLEMENTATION of BEARING FAULTS and DATA COLLECTION

To evaluate the proposed fault classification framework comprehensively, this study utilizes two distinct vibration signal datasets. The following subsections detail the data acquisition setups, fault implementations, and characteristics of these benchmark datasets, providing the necessary context for subsequent feature extraction and classification experiments.

A. The Case Western Reserve University benchmark dataset

The Case Western Reserve University (CWRU) Bearing Data Center [34] has become a standard reference for developing and evaluating bearing fault diagnosis methods. Although many studies have utilized the CWRU dataset for classification purposes [14], most of them are limited in scope—either focusing on a narrow set of fault types or reporting overly optimistic results that may not reflect real diagnostic challenges. Furthermore, certain fault conditions in the dataset—due to signal complexity, overlapping characteristics, or measurement noise—are inherently difficult to classify. A comprehensive and systematic benchmark analysis that reflects these challenges is still lacking. While an important attempt is made in [31], their study did not focus specifically on classification tasks or include a wide range of methods. To address this gap, the present study proposes a focused benchmark analysis using only the drive-end bearing data from the CWRU dataset. This dataset includes an induction motor under loads from 0-hp to 3-hp. The data were recorded at 12 kHz and 48 kHz sampling frequencies. In addition, fault diameters from 0.007 to 0.028 inches were seeded into the balls, inner race, and outer race. The application of CWRU data for validation has become a standard reference for testing fault classification algorithms. This study focuses exclusively on the drive-end bearing data to perform fault diagnosis. For experiments involving outer race faults, the defect is located in the load zone at the 6 o'clock position. Each signal is segmented into 40 non-overlapping samples, with each sample containing 2048 data points.

B. The Eskisehir Osmangazi University benchmark dataset

The three-axis vibration data used in this research were collected in Eskisehir Osmangazi University (The Center for Intelligent Systems Applications Research-CISAR) by using the testbed described in [35] and shown in Fig. 1.

A six-pole, three-phase, 1.5 kW induction motor was used in experimental data collection. The induction motor was loaded 100% with an electromagnetic powder brake (Merobel Frat 650). A National Instrument (NI) data acquisition system (NI cRio 9056) was used to record the data. Vibration signals were collected using a 3-axis vibration sensor (PCB 356a15 triaxial) via a NI 9230 module. LabVIEW tool was used for data collection. The schematic diagram of the test bed is given in Fig. 2.

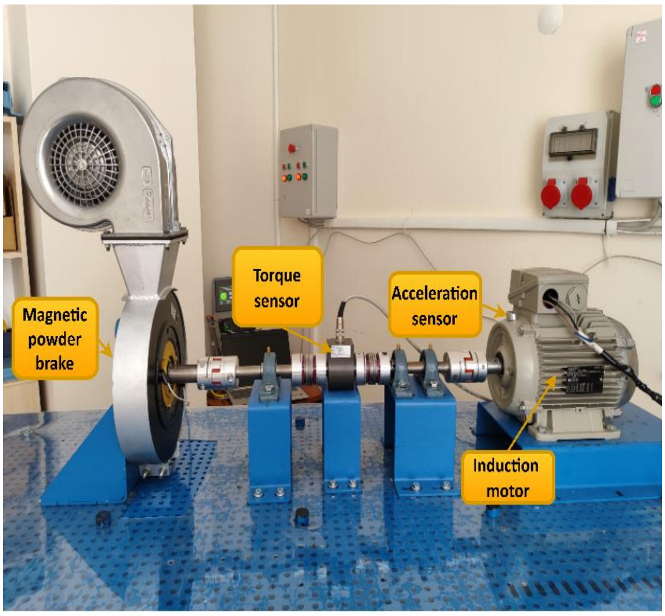


Fig. 1. The CISAR testbed.

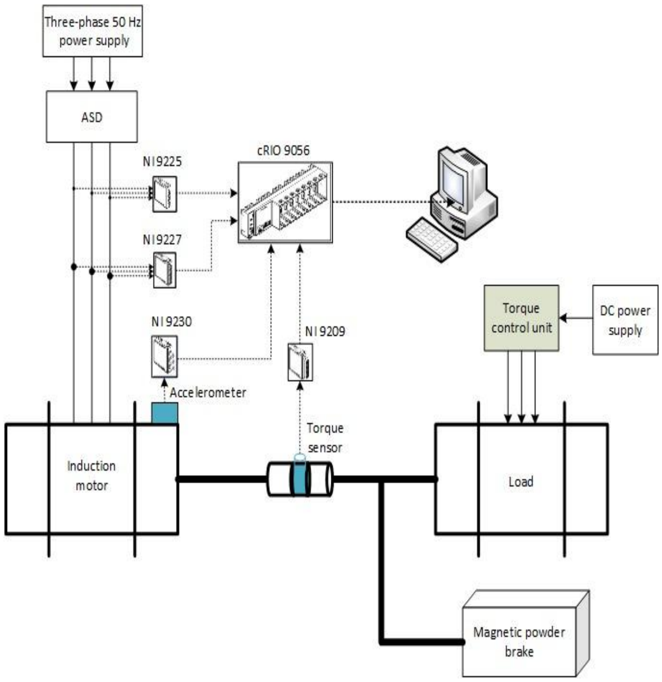


Fig. 2. Schematic diagram of testbed.

A total of four different test conditions were artificially implemented on the motor's bearing parts by using electroplating method. The four test conditions include healthy, inner race fault, outer race fault, and ball bearing fault types. The faults on bearing parts were implemented by creating holes with a diameter of 2.5 mm on the inner-race, outer-race, ball of the bearing. In each test condition, the induction motor was loaded with 100%, and the three-axis data were collected. Each bearing fault was studied independently. The three-axis vibration data of the induction motor with healthy and faulty bearings were recorded with a duration of 40 seconds and a sampling frequency of 6400 Hz. Fig. 3 shows the implemented bearing faults.

III. THE PROPOSED FAULT CLASSIFICATION FRAMEWORK WITH ENTROPY-BASED FEATURES

The classification framework consists of two main stages: feature extraction from raw vibration signals and subsequent classification using machine learning algorithms. The next sections describe these stages in detail, starting with the entropy-based methods used to derive discriminative features for bearing fault diagnosis.

A. Entropy-based feature extraction

This paper examines three entropy methods: MSE, MFE, and MPE. These methods extend sample entropy, fuzzy entropy, and permutation entropy to a multiscale framework. In the present study, these established entropy measures are not modified; rather, they are implemented using a common computational and experimental protocol to enable a controlled comparison of their feature representation capabilities for bearing fault detection.

- *MFE*: As an alternative to SamEn, fuzzy entropy (FuzzyEn) is introduced recently in [36]. Unlike SamEn, which uses the Heaviside function, FuzzyEn applies a fuzzy membership function for better continuity. It is used successfully in [37] for the detection of rolling bearing faults. Later, MFE was developed by combining FuzzyEn with a coarse-graining process to assess the self-similarity of the data [38]. However, MFEs can produce unreliable results for short-term data, highlighting the need for a more effective approach [39].
- *MSE*: Entropy-based complexity measures suggest that random sequences have the highest complexity, but in reality, neither fully regular nor entirely random sequences are the

(a)

(b)

(c)

Fig. 3. Bearing fault classes: (a) Outer race, (b) Inner race, (c) Ball defects.

most complex. Biological and physiological time series often have patterns on multiple scales. MSE, introduced in [40], captures these patterns, showing a higher complexity for correlated data and a lower complexity for predictable or random signals. MSE is widely used in time series analysis with various entropy measures [41], including Sample Entropy in [40].

- *MPE*: MPE was introduced to measure the complexity of nonlinear signals. Similar to MSE, its calculation has two steps: first, the original time series is transformed into a multiscale time series using a nonoverlapping window; then, permutation entropy is computed from the multiscale time series [42].

In this study, EntropyHub [43] in Python was used to implement these entropy-based feature extraction methods. One of the key advantages of EntropyHub is its ability to specify numerous parameters for entropy calculation through optional keyword function arguments. Drawing conclusions from entropy values is valid only when the chosen parameters accurately capture the underlying dynamics of the data.

B. The classification techniques

To ensure a fair comparison with the study of Ref. [14], the same six classification techniques were used. Each classification algorithm was implemented using the scikit-learn library for Python [44]. These algorithms are evaluated: k-Nearest Neighbors (k-NN), Linear Support Vector Machine (LSVM) with a linear kernel, Logistic regression (LR), Back-propagation neural network (BPNN), Extreme learning machine (ELM), and Softmax regression (SR).

- k-Nearest Neighbors (k-NN): The k-NN classification algorithm represents each training sample in a D-dimensional space based on its D features. A new sample is mapped to the same space, and its classification is determined by identifying its k-nearest neighbors, typically using the Euclidean distance [45], [46]. The class of these neighbors is then tallied, and the majority vote determines the classification of the new sample. In this study, the traditional k-NN rule is adapted for fault classification, where a new object is classified by analyzing its proximity to the nearest training samples in the feature space.
- *Support vector machine (SVM)*: SVMs have become a powerful tool for solving classification and regression problems. In classification tasks, SVMs aim to identify an optimal hyperplane that maximizes the separation between different classes. Conversely support vector regression seeks to approximate a mapping function from an input domain to a continuous output space while preserving the key characteristics of the maximal margin algorithm. More broadly, classification can be viewed as making predictions with binary outcomes [47], [48]. SVM classifiers can be categorized as either linear or non-linear. Scikit-learn provides three options for SVM classification, with the choice depending on factors such as the number of classes, the type of kernel function, and other model-specific hyperparameters.
- *Logistic regression (LR)*: Logistic regression, also known as maximum-entropy classification or a log-linear classifier, is a widely used supervised learning algorithm applied in various fields, including energy efficiency control and fault diagnostics [49]. LR can be implemented using three approaches: Binary, One-vs-Rest, or Multinomial. The analyzed dataset contained only two possible outcomes: faulty or healthy. In Scikit-Learn, Logistic Regression supports six solvers: (*i*) 'liblinear', (*ii*) 'lbfgs', (*iii*) 'sag', (*iv*) 'saga', (*v*) 'newton-cg', and (*vi*) 'newton-cholesky', each offering distinct advantages depending on dataset characteristics such as dimensionality, scaling, categorical features, and memory constraints.
- *Backpropagation neural network (BPNN)*: Backpropagation neural networks [50] are a type of artificial neural network that utilize multilayer feedforward architectures for computation. When used for classification, they typically consist of input, hidden, and output layers.
- *Extreme learning machine (ELM)*: Extreme learning machine, proposed in [51] is a fast and efficient single-layer feed-forward neural network algorithm [52]. Unlike traditional neural networks, ELM does not require iterative learning, which helps avoid issues such as initialization of random parameters and getting stuck in local minima. It randomly assigns input weights and biases to the hidden layer, while the output weights are determined using the least square solution. This makes ELM computationally efficient and easy to train. Because of its speed and simplicity, it has gained attention from researchers and has been applied in areas such as face recognition, time-series forecasting, and fault detection [53]. ELM supports various activation functions, including 'tanh', 'sine', 'tribas', 'inverse tribas', 'sigmoid', 'hard limit', 'soft limit', 'Gaussian', 'multiquadric', and 'inverse multiquadric'.
- *Softmax regression (SR)*: Softmax regression, an extension of logistic regression, is generally designed for multi-class classification problems where the classes are mutually exclusive. Although typically used to handle multiple classes efficiently, in this study, the softmax regression framework is adapted and applied to our binary diagnostic task (healthy vs. faulty). This allows for evaluating its feature mapping and probabilistic decision boundary capabilities in a binary context. Because of its computational simplicity and effectiveness, it has been widely applied in various practical domains [54].

IV. COMPARATIVE EVALUATION OF CLASSIFIERS and FEATURE SCALING SENSITIVITY

To assess the practical effectiveness of the proposed framework, a thorough comparative evaluation was performed across multiple classifiers and feature extraction methods. The following analysis systematically examines classification accuracy, robustness under varying conditions, and the impact of feature scaling parameters, with detailed insights supported by the empirical results presented in Tables 1-3, along with Fig. 4.

A. Hyperparameter configuration strategy

To ensure a fair and reproducible comparison across all classification models, hyperparameters were selected through an empirical validation-based search within predefined and algorithm-specific candidate ranges. Rather than performing an exhaustive grid search over all possible parameter combinations, candidate configurations were evaluated using stratified 10-fold cross-validation on the training data. The selection criterion was primarily the mean validation accuracy, while consistency across folds and model complexity were also considered to avoid overfitting. The final configurations were selected based on this comparative evaluation and are reported below. The candidate settings evaluated for each classifier and the corresponding selected configurations are detailed below; the search was conducted independently for each classifier rather than as an exhaustive Cartesian-product search across all algorithms.

- k-NN: The number of neighbors (k) was evaluated for values [1, 20], and the *n_neighbors* were determined to be 5.
- SVM: Multiple kernel functions (linear, polynomial, RBF, and sigmoid) were tested. The optimal model was found to use a linear kernel. Consequently, these hyperparameters were selected based on the dataset, with a *random_state* value of 40 to ensure reproducibility.
- LR: In this study, among the six solver options provided in Scikit-Learn, 'liblinear' was selected in combination with an L1 penalty due to its suitability for small datasets and sparsity promotion. To ensure reproducibility, the model was implemented with *random_state*=0.
- BPNN: In this study, the model is configured with 'lbfgs' as the solver, α = 1e-5 for regularization, and hidden layer sizes of (35, 25). The 'tanh' activation function is used for the hidden layers, while the output layer retains a linear activation function. In addition, a *random_state* of 25 is set to ensure reproducibility.

Table 1. Benchmark comparison of class-wise classification accuracies for 0.007 and 0.014 in. fault width based on the CWRU bearing dataset.

Fault width (in.)	Method	Classifier	Load/Speed (hp) 0 Avg. Acc. (F1 score)	0 Std	1 Avg. Acc. (F1 score)	1 Std	2 Avg. Acc. (F1 score)	2 Std	3 Avg. Acc. (F1 score)	3 Std	Mean
0.007	MSE	k-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	98.75 (94.67)	3.75	100 (100)	0	100 (100)	0	99.69
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	99.69	0.63	100	0	100	0	
	MPE	k-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MFE	k-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
0.014	MSE	k-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MPE	k-NN	98.75 (98.55)	3.75	100 (100)	0	100 (100)	0	100 (100)	0	99.69
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	98.75 (98.55)	3.75	100 (100)	0	100 (100)	0	100 (100)	0	99.69
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	99.58	1.25	100	0	100	0	100	0	
	MFE	k-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	

Table 2. Benchmark comparison of class-wise classification accuracies for 0.021 and 0.028 in. fault width based on the CWRU bearing dataset.

Fault width (in.)	Method	Classifier	Load/Speed (hp) 0 Avg. Acc. (F1 score)	0 Std	1 Avg. Acc. (F1 score)	1 Std	2 Avg. Acc. (F1 score)	2 Std	3 Avg. Acc. (F1 score)	3 Std	Mean
0.021	MSE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MPE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MFE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
0.028	MSE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MPE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	
	MFE	*k*-NN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SVM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		LR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		BPNN	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		ELM	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		SR	100 (100)	0	100 (100)	0	100 (100)	0	100 (100)	0	100
		Mean	100	0	100	0	100	0	100	0	

- ELM: After evaluating various activation functions, the ‘sigmoid’ activation function with *n_hidden*=27 can achieve the best approximation results.
- SR: In this study, the *LogisticRegression* class from *scikit-learn* is used with the parameter *multi_class= ‘multinomial’*, which allows softmax regression. The solver *lbfgs* is chosen due to its efficiency in optimizing the likelihood function for multinomial logistic regression. Furthermore, the parameter $C = 10$ is used as the inverse of regularization strength, which means that weaker regularization is applied to prevent overfitting. This configuration allows the model to effectively learn decision boundaries for multi-class classification tasks while maintaining computational efficiency.

The performance of MSE, MFE, and MPE is influenced by their respective parameter settings. In addition to classifier settings, the performance of entropy-based feature extraction methods was also found to be highly sensitive to the choice of parameter values. Although this study does not aim to propose novel parameter tuning strategies for these entropy measures, consistent parameter settings were adopted to ensure comparability and reproducibility. For all three entropy methods, the scale factor (τ) was fixed at 12 for the main experiments. The entropy parameters were selected by considering established parameter settings in the literature, consistency with [14], and their observed performance in preliminary tests. For MPE, the embedding dimension and time delay were set to m=5 and λ=1, respectively. For MSE and MFE, the embedding dimension was set to m=2, while the tolerance threshold r was defined relative to the standard deviation of the signal, as specified in reference [14]. These settings were retained to provide a consistent and reproducible comparison among the entropy-based feature extraction methods rather than introducing method-specific parameter optimization. With the scale factor set to 12, each entropy method generates a 12-dimensional feature vector, where each dimension corresponds to the entropy value calculated at one scale. Thus, MSE, MFE, and MPE independently produce 12-dimensional feature representations for each signal segment.

To investigate the impact of the scale factor on feature quality and model accuracy, a sensitivity analysis was conducted by varying τ in the range of 6 to 18. Accordingly, $\tau = 12$ was retained for the main experiments as it lies within the empirically identified optimal interval while providing a consistent parameter setting across all three entropy methods. The classification performance of all algorithms was compared across this range to identify the empirically optimal scale interval. The analysis demonstrated

Table 3. Overall classification accuracy based on the CISAR bearing dataset.

		X-axis						Y-axis						Z-axis						
		Ball		Inner		Outer		Ball		Inner		Outer		Ball		Inner		Outer		
Method	Classifier	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Avg. Acc.	Std Dvd	Mean
	k-NN	100	0	100	0	100	0	97.5	7.50	100	0	100	0	97.50	7.50	100	0	100	0	99.44
	SVM	100	0	100	0	100	0	97.5	7.50	100	0	100	0	100	0	100	0	100	0	99.72
	LR	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
MSE	BPNN	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
	ELM	87.5	20.16	95	0	97.50	7.50	85	16.58	100	0	95.00	7.50	92.50	16.01	100	0	100	7.50	94.72
	SR	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
	Mean	97.92	3.36	99.17	1.67	99.58	1.25	96.67	5.26	100	0	99.17	1.25	98.33	3.92	100	0	100	1.25	
	k-NN	100	0	100	0	100	0	100	0	100	0	100	0	87.50	16.77	97.50	7.5	100	0	98.33
	SVM	100	0	100	0	100	0	100	0	100	0	100	0	95.00	10	95.00	10	100	0	98.89
	LR	100	0	100	0	100	0	100	0	100	0	100	0	95.00	10	100	0	100	0	99.44
MPE	BPNN	100	0	100	0	100	0	100	0	100	0	100	0	97.50	7.5	100	0	100	0	99.72
	ELM	100	0	100	0	97.50	7.50	87.5	16.77	92.50	22.50	95.00	7.50	67.50	25.12	95.00	10	92.50	7.5	91.94
	SR	100	0	100	0	100	0	100	0	100	0	100	0	95.00	10	100	0	100	0	99.44
	Mean	100	0	100	0	99.58	1.25	97.92	2.8	98.75	3.75	99.17	1.25	89.58	13.23	97.92	4.58	98.75	1.25	
	k-NN	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	97.50	0	99.72
	SVM	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
	LR	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
MFE	BPNN	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
	ELM	100	0	100	0	100	0	100	0	100	0	100	0	95	10	100	0	100	0	99.44
	SR	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100	0	100
	Mean	100	0	100	0	100	0	100	0	100	0	100	0	99.17	1.67	100	0	99.58	0	

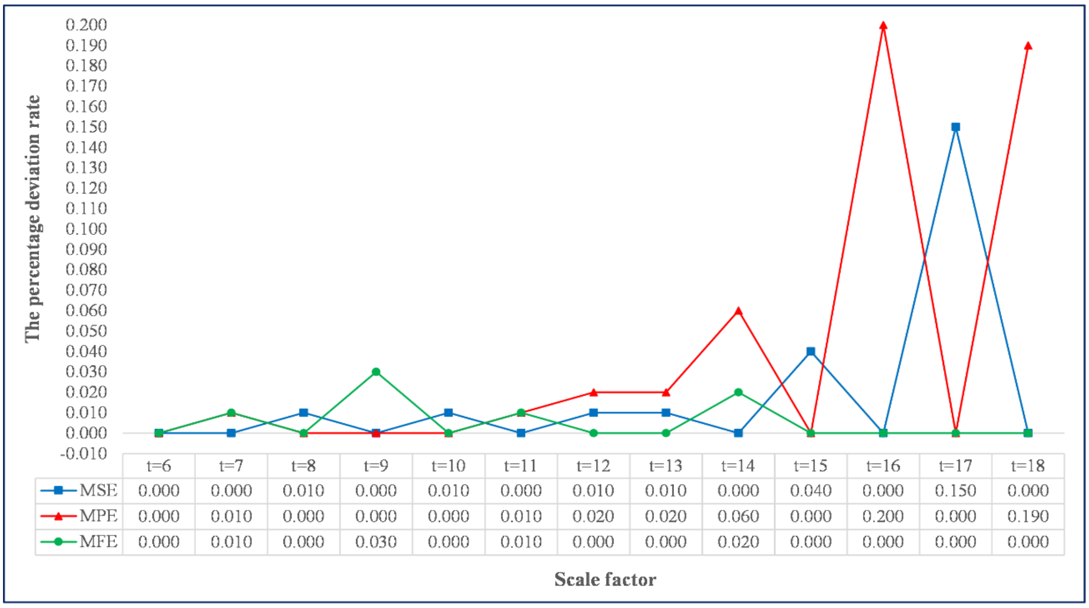


Fig. 4. Sensitivity analysis of feature extraction scale factors in fault classification performance.

that feature quality and classification accuracy were maximized when τ was in the range of 12 to 14.

B. Classification performance evaluation

The classification performance of the proposed reliability-focused framework was rigorously evaluated using two benchmark datasets: the drive-end subset of the Case Western Reserve University (CWRU) bearing data and the CISAR three-axis vibration dataset. This evaluation aimed not only to benchmark accuracy metrics but also to analyze the robustness and generalizability of the proposed entropy-based feature extraction approaches combined with six widely used classification algorithms systematically as outlined in Algorithm 1 as given in Table 4.

For each fault condition, 20 randomly selected non-overlapping signal segments were used together with 20 healthy segments, resulting in 40 observations for each binary classification task. The use of 20 samples per class was retained to maintain consistency with the benchmark configuration of [14] and to ensure a controlled and computationally manageable comparison across all entropy–classifier combinations. A fixed random seed was used for sample selection to ensure reproducibility. Thus, the classification task was formulated as binary classification (healthy vs. faulty) for each fault condition rather than as a single multi-class classification problem involving all fault types simultaneously. The same sampling strategy was applied consistently across all investigated fault conditions. Stratified 10-fold cross-validation was then applied to the resulting feature matrix to preserve the faulty-to-healthy class ratio across all folds.

For the CWRU dataset, Tables 1–2 report class-wise classification accuracies across four distinct fault types and multiple fault diameters. To ensure comparability with existing literature, particularly the benchmark study in [14], the same drive-end bearing data configuration was adopted. The results demonstrated consistent and, in many cases, highly competitive performance. Specifically, 38.18% (green-highlighted cells) of the individual classification results surpassed those reported in [14],

Table 4. Bearing fault classification algorithm using entropy features (**Algorithm 1**).

Process
Input: Bearing vibration data (0=faulty, 1=healthy)
Output: Classification accuracy for multiple entropy-based features
1. Initialize Libraries and Modules
Import pandas, numpy, scipy, sklearn, EntropyHub
2. Define *Dataset* Class
Constructor (time window, data): initialize dataset and create time-windowed samples
Method get random samples (k=20): return k random samples
3. Define *RCalculator* Class (Static Methods)
Method calculate_*r*_values (samples): r = 0.15 × SD
Method calculate_MPE (samples, M_{obj}, scale=12, method='coarse')
Method calculate_MSE (samples, r values, scale=12, method='coarse')
Method calculate_MFE (samples, r values, scale=12, method='coarse')
4. Define *ClassifierEvaluator* Class
Constructor (classifiers): initialize classifier dictionary
Method fit (X, y): scale features, perform 10-fold cross-validation, return accuracy scores
5. Main Process
Step 5.1 – Initialization:
Set random seed, define fault types, set sample length t = 2048
Step 5.2 - Process Normal Data:
Load healthy bearing data, create Dataset, sample k = 20, compute r, MPE, MSE, MFE
Export results to Excel
Step 5.3 - Process Fault Data:
for each fault type **do**
Load fault data, create Dataset, sample k = 20, compute r, MPE, MSE, MFE
Export results to Excel
end for
Step 5.4 - Prepare Classification Data:
for each entropy ∈ {MPE, MSE, MFE} **do**
Combine fault and normal features
Form feature matrix X, label vector y (0=faulty, 1=healthy)
Export to Excel
end for
Step 5.5 - Perform Classification:
Initialize classifiers: k-NN (k = 5), SVM (linear), LR, BPNN (35, 25), ELM (27), SR
for each entropy ∈ {MPE, MSE, MFE} **do**
for each classifier ∈ classifiers **do**
Perform 10-fold cross-validation
Calculate accuracy, mean, and standard deviation
Record results
end for
end for
Step 5.6 - Save Results:
Export all results to Excel

while 61.46% (yellow-highlighted cells) matched the benchmark results exactly, and only 0.35% (red-highlighted cell) showed marginally lower accuracy.

To minimize data leakage, the original CWRU vibration signals were first divided into non-overlapping segments of 2048 data points. The resulting independent segments were then used to construct the classification dataset, followed by a consistent 10-fold cross-validation procedure for all entropy–classifier combinations. The class distribution was maintained across the folds, and each segmented observation was assigned to a single fold. Thus, the cross-validation procedure was performed after signal segmentation using non-overlapping signal segments, preventing the same segmented observation from being simultaneously included in the training and test sets.

The high accuracies obtained for several classifier–entropy combinations are also consistent with [14], which used the same CWRU benchmark configuration. While MATLAB is employed in [14], the present study used Python-based implementations, providing an independent implementation of the analysis.

Among the three entropy-based feature extraction methods, MFE descriptively exhibited comparatively higher average classification accuracy on the CWRU dataset under the experimental settings considered. This comparatively higher performance may be related to MFE's ability to preserve fine-grained signal complexities across multiple scales while remaining less sensitive to noise, as also supported by the limited variation observed in the scale-factor sensitivity analysis. Overall, the average classification accuracies remained very high across the investigated classifier–entropy combinations, with the minimum mean accuracy exceeding 98%. SVM, BPNN, and SR achieved perfect classification (100% accuracy) in multiple fault scenarios. For the 0.021-inch fault diameter, all investigated classification methods achieved 100% average accuracy, whereas the 0.007-inch condition also yielded near-perfect performance, with the only notable reduction observed for the MSE–ELM combination at 1 hp (98.75%). The corresponding F1-scores reported in Tables 1 and 2 show similar trends, indicating that the high accuracy values were accompanied by consistent classification performance across the two classes. Such high levels of performance demonstrate the effectiveness of the proposed methodology under the controlled benchmark conditions considered in this study; further validation using real industrial data is required to assess its performance under practical operating conditions.

Results on the CISAR bearing dataset, summarized in Table 3, further confirmed these trends across more complex experimental measurement conditions. The three-axis vibration dataset poses additional challenges due to axis-specific noise characteristics and varying fault implementations. Nevertheless, the proposed framework maintained high classification performance across all entropy methods. Notably, for the X-axis measurements, all classifiers except for ELM achieved 100% accuracy across all entropy approaches. On the Y-axis, minor reductions in accuracy were observed for ELM as well as k-NN and SVM (when used with MSE), highlighting the nuanced challenges of axis-dependent signal characteristics.

Under these more challenging conditions, MFE also showed comparatively high average accuracy across the evaluated axes and fault types. This consistent advantage suggests that MFE may be particularly well suited for deployment in industrial monitoring systems where robustness to operational variability and noise is critical. Although ELM exhibited slightly lower average accuracies on certain axes, its computational efficiency and lightweight architecture still make it an attractive option for real-time, resource-constrained edge-computing scenarios.

An error analysis of the CISAR dataset also indicates that deviations from perfect accuracy tended to concentrate in signals with smaller fault diameters or lower signal-to-noise ratios, underscoring the inherent difficulty of detecting early-stage or subtle fault features. This observation aligns with practical diagnostic challenges, where early fault detection remains the most critical—and most difficult—aspect of predictive maintenance.

Taken together, the results indicate that the proposed framework provides consistent classification performance across the investigated datasets, fault conditions, and measurement axes. The complementary F1-score results for the CWRU dataset further support the high classification performance observed in terms of accuracy. However, these findings should be interpreted within the controlled benchmark settings and binary healthy-versus-faulty classification design considered in this study. The observed results therefore provide empirical evidence for the effectiveness of the investigated entropy–classifier combinations, while further validation on larger and heterogeneous industrial datasets is required to assess their generalization capability.

C. Sensitivity analysis of scale factor in feature extraction

The sensitivity analysis highlights the pivotal importance of selecting appropriate parameters during signal preprocessing and feature extraction. A critical aspect of the proposed framework involves understanding how parameter choices in feature extraction influence overall classification performance. In particular, the scale factor (τ) used in multiscale entropy-based feature extraction is known to affect the granularity at which temporal signal dynamics are captured. A poorly chosen scale factor can obscure important fault signatures or introduce artifacts, ultimately degrading classification accuracy. Although other entropy parameters, including the embedding dimension, time delay, and tolerance threshold, may also affect the extracted feature representations and classification performance, these parameters were kept fixed in the present study based on established settings and the parameterization reported in [14]. The sensitivity analysis was therefore specifically designed to investigate the effect of the scale factor while maintaining the other entropy parameters constant. A systematic sensitivity analysis of these remaining parameters was beyond the scope of the present study and is considered an important direction for future research.

To systematically investigate this effect, a detailed sensitivity analysis was conducted in which τ was varied from 6 to 18. For each value of τ, features were extracted using the three entropy methods—MSE, MPE, and MFE—and classification accuracy was computed for each of the six classifiers. The analysis was designed to quantify how variations in τ influence the stability and robustness of the feature representations under different fault types and operating conditions.

Fig. 4 provides a comprehensive visualization of the sensitivity analysis. The y-axis represents the total deviation of average classification accuracy from the ideal benchmark of 100%, while the x-axis corresponds to the scale factor. Lower deviation values indicate that a feature extraction method yields stable and reliable representations that generalize well across fault types and operating conditions.

Several important trends emerge from this analysis. For MSE and MPE, accuracy deviations increase noticeably as the scale factor rises beyond approximately 11. This trend suggests that larger scale factors may over-smooth the temporal dynamics of vibration signals, effectively filtering out subtle fault signatures necessary for accurate classification. Such over-smoothing reduces the discriminative power of the extracted features, leading to higher classification errors. Conversely, very low scale factors risk capturing only fine-grained noise rather than meaningful signal structure.

By contrast, MFE showed comparatively lower variation in the deviation values across the investigated scale-factor range. Its deviation curve remains relatively flat across the entire range of τ, with only minor local variations observed at values such as 7, 9, 11, and 14. This observed pattern may indicate that MFE provides greater tolerance to scale-factor variation under the experimental conditions considered. In practical terms, this property implies that MFE-based features require less aggressive parameter tuning, making them especially suitable for deployment in real-world monitoring systems where operating conditions and signal characteristics can vary unpredictably.

The results of this sensitivity analysis also carry significant implications for predictive maintenance strategies in industrial settings. Specifically, they underscore that reliable fault classification systems must include careful calibration of feature extraction parameters to ensure consistent performance. For example, while MSE and MPE can achieve high classification accuracy when τ is optimally selected (e.g., in the 12–14 range identified in this study), their sensitivity to parameter changes necessitates more careful, potentially adaptive tuning in live deployments. The comparatively lower sensitivity observed for MFE may be advantageous in settings where extensive parameter re-tuning is undesirable; however, this potential practical implication requires further validation under real operating conditions.

Moreover, these findings suggest a need for developing adaptive or data-driven scale-factor selection techniques as a promising future direction. Such methods could dynamically optimize τ based on real-time signal statistics, further enhancing system robustness and diagnostic reliability.

In conclusion, the scale-factor sensitivity analysis conducted in this work provides essential insights into the interplay between feature extraction parameterization and classification performance. It emphasizes the critical importance of parameter selection in signal preprocessing pipelines and provides descriptive comparative insights into the scale-factor sensitivity of MFE in industrial fault diagnosis applications. These insights not only validate the methodological choices made in this study but also offer practical guidelines for deploying robust, real-time bearing fault classification systems in challenging operating environments.

V. DISCUSSION and CONCLUSION

This study presents a comprehensive, reliability-focused framework for bearing fault diagnosis that integrates multiscale entropy-based feature extraction with six widely used classification algorithms. Through systematic benchmarking on the Case Western Reserve University (CWRU) and CISAR three-axis vibration datasets, the proposed framework demonstrated high classification accuracy and consistent performance across the investigated experimental conditions, while also providing insights into the comparative behavior of the entropy-based feature extraction methods.

- *Key Findings and Contributions*: A central contribution of this work is the unified benchmark that evaluates three entropy-based feature extraction methods—MSE, MPE, and MFE—in combination with six classification techniques. The experiments descriptively indicated that MFE frequently achieved comparatively higher classification accuracy across the datasets and classifiers considered. This observed performance pattern may be associated with MFE's ability to maintain meaningful signal complexity representations while demonstrating resilience to noise and operational variability.

  The high accuracies observed in the CWRU experiments should be interpreted in the context of the controlled benchmark setting and the binary healthy-versus-faulty classification design. The consistency of many results with [14] further indicates that these high-performance levels are not solely attributable to the implementation used in this study. However, the more heterogeneous CISAR measurements demonstrate that classification performance can vary with signal characteristics and measurement axes. The relatively consistent performance of MFE across these conditions suggests that it may provide a more robust entropy representation under the experimental conditions considered.

  The scale-factor analysis also indicates that parameter robustness is an important consideration when selecting entropy-based representations. The comparatively greater sensitivity of MSE and MPE suggests that their discriminative capability may depend more strongly on the scale at which signal dynamics are represented, whereas the lower sensitivity observed for MFE indicates greater tolerance to scale-factor variation. This distinction is particularly relevant for practical monitoring systems, where operating conditions may change and extensive parameter re-tuning may not be desirable. The identified 12–14 interval therefore provides an empirical reference for similar experimental settings rather than a universal optimal range.

  The performance differences observed among MSE, MFE, and MPE should be interpreted as descriptive comparative findings within the experimental settings considered. The reported mean accuracy and standard deviation values indicate relative differences in classification performance and variability; however, no formal statistical significance analysis was conducted. Accordingly, observations referring to comparatively higher or more stable performance should not be interpreted as claims of statistically significant differences. A dedicated statistical analysis based on repeated experimental trials and a larger set of independent observations is an important direction for future research.

- *Practical Implications*: From an industrial perspective, the main value of the proposed framework lies in combining entropy-based representations with computationally lightweight classifiers within a unified diagnostic workflow. The comparatively stable behavior of MFE and the performance of lightweight classifiers such as ELM indicate the potential for developing monitoring systems that require limited feature engineering and can operate under constrained computational resources. The identified scale-factor range may also serve as an initial parameter-selection guideline for similar bearing-monitoring applications. However, these implications represent potential applicability rather than validated real-time industrial deployment, since execution time, memory requirements, and performance on live industrial data were not evaluated in this study.

  Furthermore, the consistent performance observed across the two evaluated datasets indicates the potential applicability of the framework to a broader range of rotating machinery. However, such applicability should be further validated using heterogeneous datasets and operational industrial data before broader deployment in sectors such as manufacturing, transportation, and energy.

- *Limitations and Future Work*: Despite these promising results, several limitations warrant discussion. First, while the two datasets used in this study offer diversity in terms of operating conditions and fault types, they remain controlled environments with artificially induced faults. Real-world industrial deployments may face additional challenges such as evolving fault characteristics, variable loading conditions, and non-stationary noise. As such, further validation of operational data from live industrial systems is essential. In addition, the CWRU and CISAR datasets were evaluated independently rather than through direct cross-dataset validation. Although consistent performance trends were observed across the two datasets, these results do not constitute a direct assessment of cross-dataset generalization. Future research should therefore evaluate the proposed framework using cross-dataset validation and additional heterogeneous datasets, particularly independently collected industrial data, to further assess its generalization capability. Furthermore, the number of observations used for each binary classification task was relatively limited. Although stratified 10-fold cross-validation was employed to preserve class balance, the sample size may affect the statistical stability and generalizability of the reported performance estimates. Future studies should therefore consider larger numbers of independently sampled observations and repeated cross-validation procedures.

  A second limitation concerns the sensitivity analysis, which was restricted to the scale factor over the predefined range of 6–18 while the remaining entropy parameters were fixed. Future research could explore adaptive and data-driven scale-factor selection methods that dynamically optimize parameter choices based on real-time signal characteristics, potentially further improving the robustness of the proposed framework.

  Another limitation is that the study benchmarked six classification algorithms without considering advanced deep learning architectures. Future investigations could explore the integration of entropy-based features with models such as convolutional neural networks and transformer-based architectures to capture more complex fault patterns, particularly under highly non-stationary or noisy operating conditions.

  Finally, the computational cost, feature-extraction runtime, inference time, and memory requirements of the proposed framework were not quantitatively evaluated. Although the classification results indicate the potential applicability of the proposed framework to practical monitoring applications, dedicated runtime and resource analyses are required before its suitability for real-time or edge deployment can be established. Future studies should therefore investigate feature-extraction and inference times, memory usage, and computational complexity of the proposed framework under representative edge-computing hardware constraints.

Overall, the findings demonstrate that multiscale entropy-based feature extraction combined with conventional classifiers can provide reliable bearing fault discrimination under the benchmark conditions investigated. MFE showed comparatively higher and more stable performance across the evaluated settings, while the scale-factor analysis highlighted the importance of parameter robustness. The study therefore provides a consistent comparative benchmark and practical parameter-selection insights, while the identified limitations highlight the need for validation on larger, heterogeneous, and independently collected industrial datasets and for dedicated computational-cost analysis before real-world deployment.

AUTHOR STATEMENT

Plagiarism Check—The article has been scanned with iThenticate and found to be compliant with the journal's plagiarism policy.
Conflict of Interest—There is no conflict of interest with any person/organization.
Ethics Committee Approval—Ethics committee approval is not required for this article.
Use of Artificial Intelligence Tools—In this study, ChatGPT-X was used for grammar and spell checking. All content reflects the original contribution of the authors.
Funding—No institutional/financial support was received for this study.
Data availability—The data supporting the findings of this study were generated by the author(s) and may be obtained from the author(s) upon request.
CRediT Author Contribution—Conceptualization, Methodology, Software, Formal analysis, Investigation, Formal analysis, Validation, Visualization, Writing - original draft, review & editing. (**Derya Deliktaş**); Conceptualization, Methodology, Software, Formal analysis, Validation, Visualization, Writing-review & editing. (**Gurkan Ozturk**); Conceptualization, Methodology, Validation, Visualization, Writing-review & editing. (**Abdurrahman Ünsal**); Conceptualization, Methodology, Investigation, Visualization, Writing-review & editing. (**Ozden Ustun**).

REFERENCES

[1] K. C. Rath, A. Khang, D. Roy, "The role of internet of things (IoT) technology in industry 4.0 economy", *Advanced IoT Technologies and Applications in the Industry 4.0 Digital Economy*, 1st Edition, A. Khang, V. Abdullayev, V. Hahanov, V. Shah, Eds. Boca Raton, Florida, USA, *CRC Press*, 2024, Ch. 1. https://doi.org/10.1201/9781003434269.

[2] N. D. Thuan, H. S. Hong, "HUST bearing: a practical dataset for ball bearing fault diagnosis", *BMC Research Notes*, 16(1), 138, 2023. https://doi.org/10.1186/s13104-023-06400-4.

[3] C. Wang, J. Yang, H. Jie, Z. Zhao, W. Wang, "An energy-efficient mechanical fault diagnosis method based on neural dynamics-inspired metric SpikingFormer for insufficient samples in industrial internet of things", *IEEE Internet of Things Journal*, 12(1), 1081-1097, 2025. https://doi.org/10.1109/JIOT.2024.3476034.

[4] Y. Shang, X. Tang, G. Zhao, P. Jiang, T. R. Lin, "A remaining life prediction of rolling element bearings based on a bidirectional gate recurrent unit and convolution neural network", *Measurement*, 202, 111893, 2022. https://doi.org/10.1016/j.measurement.2022.111893.

[5] Q. Ni, J. C. Ji, K. Feng, "Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network", *IEEE Transactions on Industrial Informatics*, 19(2), 1301-1311, 2023. https://doi.org/10.1109/TII.2022.3169465.

[6] J. Guo, Z. Li, M. Li, "A review on prognostics methods for engineering systems", *IEEE Transactions on Reliability*, 69(3), 1110–1129, 2020. https://doi.org/10.1109/TR.2019.2957965.

[7] M. A. Abbasi, S. P. Huang, A. S. Khan, "Fault detection and classification of motor bearings under multiple operating conditions", *ISA Transactions*, 156, 61–69, 2025. https://doi.org/10.1016/j.isatra.2024.11.008.

[8] J. Jiao, M. Zhao, J. Lin, K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis", *Neurocomputing*, 417, 36–63, 2020. https://doi.org/10.1016/j.neucom.2020.07.088.

[9] Y. Wang, J. Wu, Y. Cheng, J. Wang, K. Hu, "Memory-enhanced hybrid deep learning networks for remaining useful life prognostics of mechanical equipment", *Measurement*, 187, 110354, 2022. https://doi.org/10.1016/j.measurement.2021.110354.

[10] M. A. Abbasi, S. Huang, "Development of a fault diagnostics and tolerance system: An application to continuous stirred tank reactor", *Measurement Science and Technology*, 35(6), 066203, 2024. https://doi.org/10.1088/1361-6501/ad2da7.

[11] Q. Ni, J. C. Ji, K. Feng, Y. Zhang, D. Lin, J. Zheng, "Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit", *Reliability Engineering & System Safety*, 242, 109753, 2024. https://doi.org/10.1016/j.ress.2023.109753.

[12] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap", *Mechanical Systems and Signal Processing*, 138, 106587, 2020. https://doi.org/10.1016/j.ymssp.2019.106587.

[13] S. Chikkam, S. Singh, "Condition monitoring and fault diagnosis of induction motor using DWT and ANN", *Arabian Journal for Science and Engineering*, 48(5), 6237–6252, 2023. https://doi.org/10.1007/s13369-022-07294-3.

[14] Y. Li, X. Wang, S. Si, S. Huang, "Entropy based fault classification using the Case Western Reserve University data: A benchmark study", *IEEE Transactions on Reliability*, 69(2), 754–767, 2020. https://doi.org/10.1109/TR.2019.2896240.

[15] P. Wang, H. Xiong, H. He, "Bearing fault diagnosis under various conditions using an incremental learning-based multi-task shared classifier", *Knowledge-Based Systems*, 266, 110395, 2023. https://doi.org/10.1016/j.knosys.2023.110395.

[16] P. K. Kankar, S. C. Sharma, S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods", *Expert Systems with Applications*, 38(3), 1876–1886, 2011. https://doi.org/10.1016/j.eswa.2010.07.119.

[17] L. Frosini, C. Harlisca, L. Szabo, "Induction machine bearing fault detection by means of statistical processing of the stray flux measurement", *IEEE Transactions on Industrial Electronics*, 62(3), 1846–1854, 2015. https://doi.org/10.1109/TIE.2014.2361115.

[18] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, "Deep learning and its applications to machine health monitoring", *Mechanical Systems and Signal Processing*, 115, 213–237, 2019. https://doi.org/10.1016/j.ymssp.2018.05.050.

[19] Y. Yang, M. M. M. Haque, D. Bai, W. Tang, "Fault diagnosis of electric motors using deep learning algorithms and its application: A review", *Energies*, 14(21), 7017, 2021. https://doi.org/10.3390/en14217017.

[20] J. Yin, G. Cen, "Intelligent motor bearing fault diagnosis using channel attention-based CNN", *World Electric Vehicle Journal*, 13(11), 208, 2022. https://doi.org/10.3390/wevj13110208.

[21] G. Li, Y. Chen, W. Wang, Y. Wu, R. Liu, "Automatic Transmission Bearing Fault Diagnosis Based on Comprehensive Index Method and Convolutional Neural Network", *World Electric Vehicle Journal*, 13(10), 184, 2022. https://doi.org/10.3390/wevj13100184.

[22] S. Zhang, S. B. Zhang, B. Wang, T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review", *IEEE Access*, 8, 29857–29881, 2020. https://doi.org/10.1109/ACCESS.2020.2972859.

[23] Z. Liu, P. Zhang, Y. Yu, M. Li, Z. Zeng, "A novel fault diagnosis model of rolling bearing under variable working conditions based on attention mechanism and domain adversarial neural network", *Journal of Mechanical Science and Technology*, 38(3), 1101–1111, 2024. https://doi.org/10.1007/s12206-024-1208-8

[24] L. H. Wang, X. P. Zhao, J. X. Wu, Y. Y. Xie, Y. H. Zhang, "Motor fault diagnosis based on short-time Fourier transform and convolutional neural network", *Chinese Journal of Mechanical Engineering*, 30(6), 1357–1368, 2017. https://doi.org/10.1007/s10033-017-0190-5.

[25] C. Deng, Z. Deng, S. Lu, M. He, J. Miao, Y. Peng, "Fault diagnosis method for imbalanced data based on multi-signal fusion and improved deep convolution generative adversarial network", *Sensors*, 23(5), 2542, 2023. https://doi.org/10.3390/s23052542.

[26] F. Chen, X. Zhou, B. Xu, Z. Yang, Z. Qu, "Instantaneous square current signal analysis for motors using vision transformer for the fault diagnosis of rolling bearings", *Applied Sciences*, 13(16), 9349, 2023. https://doi.org/10.3390/app13169349.

[27] G. Niu, X. Dong, Y. Chen, "Motor fault diagnostics based on current signatures: A review", *IEEE Transactions on Instrumentation and Measurement*, 72, 2023. https://doi.org/10.1109/TIM.2023.3285999.

[28] Z. Wang, C. Guan, S. Shi, G. Zhang, X. Gu, "Motor Bearing Fault Diagnosis Based on Current Signal Using Time–Frequency Channel Attention", *World Electric Vehicle Journal*, 15(7), 281, 2024. https://doi.org/10.3390/wevj15070281.

[29] M. A. Abbasi, S. Huang, A. S. Khan, "Hybrid fault diagnostics and prognostics system development for motor bearings health monitoring", *Quality Engineering*, 37(2), 292–307, 2025. https://doi.org/10.1080/08982112.2024.2385920.

[30] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests", *IEEE International Conference on Prognostics and Health Management (PHM'12)*, Denver, Colorado, USA, June 2012, https://hal.science/hal-00719503v1.

[31] W. A. Smith, R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study", *Mechanical Systems and Signal Processing*, 64-65, 100–131, 2015. https://doi.org/10.1016/j.ymssp.2015.04.021.

[32] Y. Su, L. Shi, K. Zhou, G. Bai, Z. Wang, "Knowledge-informed deep networks for robust fault diagnosis of rolling bearings", *Reliability Engineering & System Safety*, 244, 109863, 2024. https://doi.org/10.1016/j.ress.2023.109863.

[33] V. Pandhare, J. Singh, J. Lee, "Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features", Proceedings of the Prognostics and System Health Management Conference (PHM-Paris), Paris, France, 02-05 May 2019, 320–326. https://doi.org/10.1109/PHM-Paris.2019.00061.

[34] Case Western Reserve University, "Bearing data center test seeded fault test data", https://engineering.case.edu/bearingdatacenter/download-data-file (14.09.2026).

[35] E. Irgat, E. Çinar, A. Ünsal, A. Yazıcı, "An IoT-based monitoring system for induction motor faults utilizing deep learning models", *Journal of Vibration Engineering & Technologies*,11(7), 3579–3589, 2023. https://doi.org/10.1007/s42417-022-00769-5.

[36] W. Chen, J. Zhuang, W. Yu, Z. Wang, "Measuring complexity using fuzzyen, apen, and sampen", *Medical Engineering & Physics*, 31(1), 61–68, 2009. https://doi.org/10.1016/j.medengphy.2008.04.005.

[37] C. S. Park, Y. C. Choi, Y. H. Kim, "Early fault detection in automotive ball bearings using the minimum variance cepstrum", *Mechanical Systems and Signal Processing*, 38(2), 534–548, 2013. https://doi.org/10.1016/j.ymssp.2013.02.017

[38] J. Zheng, J. Cheng, Y. Yang, S. Luo, "A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination", *Mechanism and Machine Theory*, 78, 187–200, 2014. https://doi.org/10.1016/j.mechmachtheory.2014.03.014.

[39] Y. Li, M. Xu, R. Wang, W. Huang, "A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy", *Journal of Sound and Vibration*, 360, 277–299, 2016. https://doi.org/10.1016/j.jsv.2015.09.016.

[40] M. Costa, A. L. Goldberger, C. K. Peng, "Multiscale entropy analysis of complex physiologic time series", *Physical Review Letters*, 89(6), 068102, 2002. https://doi.org/10.1103/PhysRevLett.89.068102.

[41] F. C. Morabito, D. Labate, F. L. Foresta, A. Bramanti, G. Morabito, I. Palamara, "Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG", *Entropy*, 14(7), 1186–1202, 2012. https://doi.org/10.3390/e14071186.

[42] Y. Li, Q. Gao, B. Miao, W. Zhang, J. Liu, Y. Zhu, "Application of the refined multiscale permutation entropy method to fault detection of rolling bearing", *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 43(5), 280, 2021. https://doi.org/10.1007/s40430-021-02986-7.

[43] M. W. Flood, B. Grimm, "Entropyhub: An opensource toolkit for entropic time series analysis", *PloS One*, 16(11), e0259448, 2021. https://doi.org/10.1371/journal.pone.0259448.

[44] F. Pedregosa et al., "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, 12, 2825–2830, 2011.

[45] J. Wang, P. Neskovic, L. N. Cooper, "Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence", *Pattern Recognition*, 39(3), 417–423, 2006. https://doi.org/10.1016/j.patcog.2005.08.009.

[46] S. R. Madeti, S. N. Singh, "Modeling of PV system based on experimental data for fault detection using kNN method", *Solar Energy*, 173, 139–151, 2018. https://doi.org/10.1016/j.solener.2018.07.038.

[47] N. Cristianini, J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods", Cambridge University Press, 2000. https://doi.org/10.1017/CBO9780511801389.

[48] U. B. Parikh, B. Das, R. Maheshwari, "Fault classification technique for series compensated transmission line using support vector machine", *International Journal of Electrical Power & Energy Systems*, 32(6), 629–636, 2010. https://doi.org/10.1016/j.ijepes.2009.11.020.

[49] P. Barandier, M. Mendes, A. J. M. Cardoso, "Comparative analysis of four classification algorithms for fault detection of heat pumps", *Energy and Buildings*, 316, 114342, 2024. https://doi.org/10.1016/j.enbuild.2024.114342.

[50] Y. Li, W. Dong, Q. Yang, J. Zhao, L. Liu, S. Feng, "An automatic impedance matching method based on the feedforward-backpropagation neural network for a WPT system", *IEEE Transactions on Industrial Electronics*, 66(5), 3963–3972, 2019. https://doi.org/10.1109/TIE.2018.2835410.

[51] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: theory and applications", *Neurocomputing*, 70(1-3), 489–501, 2006. https://doi.org/10.1016/j.neucom.2005.12.126.

[52] A. V. N. Reddy, C. P. Krishna, P. K. Mallick, "An image classification framework exploring the capabilities of extreme learning machines and artificial bee colony", *Neural Computing and Applications*, 32(8), 3079–3099, 2020. https://doi.org/10.1007/s00521-019-04385-5.

[53] X. Gao, G. Yuan, M. Zhang, "Fault detection of electric vehicle charging piles based on extreme learning machine algorithm", *Proceedings of the 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 11-13 March 2020, 849–852. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000157.

[54] M. Jiang, Y. Liang, X. Feng, X. Fan, Z. Pei, Y. Xue, R. Guan, "Text classification based on deep belief network and softmax regression", *Neural Computing and Applications*, 29(1), 61–70, 2018. https://doi.org/10.1007/s00521-016-2401-x.