INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING
ISSN: 2229-8649 e-ISSN: 2180-1606
VOLUME 23, ISSUE 1, 2026, 13419 - 13433
DOI: https://doi.org/10.15282/ijame.23.1.2026.20.1018

RESEARCH ARTICLE

Generalized bearing fault diagnosis using comparative time-frequency scalogram features and machine learning with advanced feature ranking

Vipul Dave[1]*, Vinay Vakharia[2], Pradeep Kumar Karsh[1] ,Vipul M Dabhi[3], Khyati Zalawadia[4]

[1]Department of Mechanical Engineering, Parul Institute of Engineering & Technology, Parul University, 391760, Vadodara, Gujarat, India
[2]Department of Mechanical Engineering, Pandit Deendayal Energy University, 382007 Gandhinagar, India
[3]Department of Computer Science Engineering, Gokul Global University, Sidhpur, 384151, Gujarat, India
[4]Department of Computer Engineering, Parul Institute of Engineering & Technology, Parul University, 391760, Vadodara, Gujarat, India

Abstract – Bearings are critical components of rotating machinery, ensuring smooth operation by supporting shafts and loads. Even minor defects can severely degrade performance, causing unplanned downtime and economic loss. Accurate and timely fault diagnosis is therefore essential for condition-based maintenance. Conventional time- and frequency-domain methods often fail to capture the non-stationary and transient nature of bearing vibration signals. This work proposes a time–frequency diagnostic framework using Continuous Wavelet Transform (CWT)–based scalograms and machine learning. Coiflet and Morlet wavelets are selected using the Maximum Relative Wavelet Energy (MRWE) criterion, yielding peak MRWE values of 0.1347 for the CWRU dataset and 0.0584 for the Machinery Fault Simulator (MFS) dataset, ensuring optimal fault localization. Scalograms are converted into RGB images, from which 21 texture features are extracted and ranked using the Fisher score. The method is validated on two independent datasets: the benchmark Case Western Reserve University (CWRU) data and an experimental MFS dataset using SKF 6004 bearings with induced inner-race, outer-race, and ball defects. Logistic Regression, SVM, KNN, and Bagged Tree classifiers are evaluated using tenfold cross-validation. On the CWRU dataset, KNN and Bagged Tree achieve accuracies of 98.3% with eight ranked features, while the Bagged Tree reaches 100% accuracy using all 21 features.

Article History
Received : 12 September 2025
Revised : 28 November 2025
Accepted : 08 January 2026
Published : 15 March 2026

Keywords
Bearing faults
Time-frequency analysis
Wavelet transform
KNN
Bagged Tree

1. Introduction

Rolling bearings are universal elements of rotating machinery, commonly found in automotive, aerospace, manufacturing, and energy applications, where they support shafts and transmit motion while bearing radial and axial loads. Although the design criteria are rigorous, the manufacturing process is highly precise, and the tests are stringent, bearings often do not reach the stated service life. Such premature failure may be due to many different causes, such as material fatigue, insufficient lubrication, particle contamination, improper mounting, misalignment, unforeseen load changes, or poor fit tolerances. This is not a one-off occurrence- bearing degradation can easily spread through the machine system, resulting in vibration, thermal stress, severe performance degradation, and eventually a disastrous breakdown. These impacts are directly translated into financial losses, unplanned downtime, and damage to adjacent components [1]. Thus, early and precise fault diagnosis of bearing faults has become an essential prerequisite to industrial condition-based monitoring and prediction-based maintenance programs.

Traditional bearing fault diagnosis schemes can be roughly divided into three areas: time-domain, frequency-domain, and hybrid approaches. Simple, computationally low-cost time-domain techniques, including mean, variance, RMS value, crest factor, skewness, and kurtosis, are used to calculate the statistical properties of a time-varying signal [2]. Yet they are no good at capturing transient signatures in non-stationary signals and are very sensitive to background noise. It is common to use frequency-domain techniques, typically the Fast Fourier Transform (FFT) or the Hilbert Transform, to detect characteristic fault frequencies associated with inner-race, outer-race, or rolling-element defects. Although FFT - based methods are effective for stationary, periodic signals, speed variations, modulations, or complex machine dynamics can often render the procedure less sensitive [3]. It has been proposed that hybrid signal processing methods be used to address these limitations. Intrinsic mode functions, as used by Ensemble Empirical Mode Decomposition, to extract localized patterns in a signal; instantaneous frequency, provided by Hilbert Transform and spectral analysis; impulsive faults, which are identified through spectral kurtosis; and robust detection of defects, which is achieved by probabilistic principal component analysis and statistical filtering, are examples. These methods have high diagnostic sensitivity but can be computationally expensive, produce non-unique decompositions, and are sensitive to parameter choice and noise contamination [4]. On the contrary, time-frequency analysis techniques, especially the Continuous Wavelet Transform (CWT), are better suited for non-stationary and transient vibration signals. CWT produces scalograms that can simultaneously record the time-dependent and spectral development of vibration energy, allowing short-lived defect impulses hidden in other domains to be identified [5]. Experiments have demonstrated that the wavelet scalogram's sensitivity to various fault frequencies can be improved by filtering into red, green, and blue (RGB) channels, and that each channel can accentuate a different defect mode. Here, Haar, Symlet, Coiflet, Morlet, and Daubechies (db2) wavelets are tested, and Maximum Relative Wavelet Energy (MRWE) is applied to the wavelets to select the best wavelet to

*CORRESPONDING AUTHOR | Vipul Dave | ✉ vipul.dave270188@paruluniversity.ac.in

represent faults [6]. This wavelet selection method is systematic, avoiding arbitrary choices in decomposition, and enhances diagnostic consistency.

Feature extraction and ranking are important for enhancing the quality of diagnosis. RGB scalograms should be statistically summarised to remove redundancy and sensitivity to noise. Making features as discriminative as possible with minimal computational effort: Fisher Score, the feature ranking technique that maximizes the distance between classes between healthy and faulty data. ReliefF Algorithm, where feature value is estimated using the nearest neighbor differences between a class, and is therefore strong with noisy data. Minimum Redundancy Maximum Relevance (mRMR) chooses features that are strongly correlated with the desired output but weakly correlated with one another to enhance the performance of the classifier. The Principal Component Analysis-based ranking, also called a conversion of the original feature space into orthogonal components ranked by how much they account for the signal variance; dimensional reduction with dominant information conservation [7]. When used appropriately, these ranking techniques can reduce the number of features needed without compromising classification accuracy, enabling real-time applications and reducing computational load.

In the current paper, the vibration data is represented by two datasets: the Case Western Reserve University (CWRU) benchmark, which is well known, and an experimental dataset gathered on a Machinery Fault Simulator (MFS) rig. The MFS arrangement includes a horizontal rotor mounted on SKF 6004 deep groove ball bearings and connected to a DC motor by servo means to change speed between 200 and 3000 rpm in 200 rpm steps. A systematic analysis of fault conditions, such as inner- or outer-race defects or rolling-element defects, is artificially induced. A multi-axial piezoelectric accelerometer is employed to record high-resolution vibration signals with a sensitivity of 10.33 mV/g and a sampling frequency of 12 kHz. The overall workflow of the proposed methodology is illustrated in Figure 1. The comparison between these two datasets shows that the proposed methodology is not data-dependent and yields generalizable results across laboratories and real experimental conditions. Once features have been extracted from optimized wavelet scalograms, they are ranked using the Fisher Score. Classifiers based on machine learning, such as Logistic Regression, Support Vector Machine (SVM), Fine K-Nearest Neighbor (KNN), and Bagged Tree, are applied to diagnose and classify bearing faults [8].

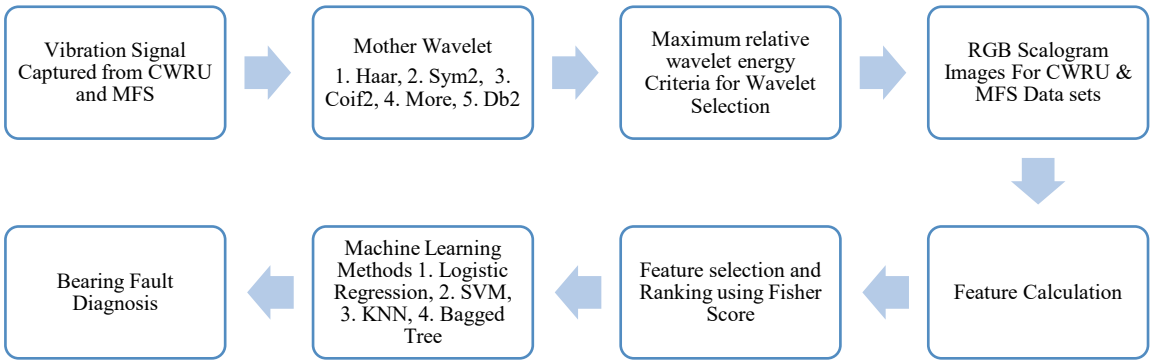

Figure 1. Flow chart of the process

The comparative findings affirm that the combination of CWT-based scalogram features, optimal wavelet selection via MRWE, and advanced feature ranking methods yields higher diagnostic accuracy than conventional vibration analysis. The originality of the current work lies not only in the hybrid approach but also in the dual validation on CWRU and MFS data. This ensures the scalability and industry applicability of the proposed fault diagnosis framework, which can be quickly scaled and adjusted to different fault rates, fault severities, and levels of noise with no manual retuning. The system can achieve very high fault-detection rates at low computational cost by selecting the most discriminative features and using powerful machine learning models [9]. Such an approach will enhance predictive maintenance, increase the reliability of machinery, and provide a solid basis for future investigations into automated health monitoring systems for rotating machines.

2. Materials and Methods

The methodology describes the Signal processing techniques used for analysis, as well as the selection of the mother wavelet based on the maximum relative wavelet energy criterion. Feature ranking and the use of different machine learning algorithms for analysis are also discussed in detail.

2.1 Signal Processing Technique and Wavelet Selection

The Continuous Wavelet Transform is an effective signal processing method that provides time-frequency localization that cannot be achieved by other transformations, such as the Short-Time Fourier Transform (STFT) or the classical Fourier Transform (FT). In contrast to FT, which gives frequency information but no time resolution, or STFT, which gives joint resolution (depending on a fixed window size), CWT uses variable-sized windows, which resize depending on the frequency content under analysis [10]. In particular, the CWT offers both high frequency resolution at low frequencies and better time resolution at high frequencies, which are suited to non-stationary or transient signal analysis. In CWT, local and global features are both revealed at the same time as the original signal is projected onto a family of wavelet functions produced by scaling (dilation) and translating a single mother wavelet. This multiresolution scheme

allows one to detect minute variations in patterns of signals, otherwise obscured by fixed-resolution schemes. On a conceptual level, STFT applies Fourier analysis to windowed segments, whereas CWT does not require explicit Fourier analysis of window functions (single-peak spectral ambiguities are avoided), and instead generates a continuous representation that smoothly changes with scale [11]. In addition to this, the frame size of the CWT naturally changes with scale, so small wavelets are employed to encode high-frequency information with a fine time localization, and large wavelets encode low-frequency trends with a sharp frequency localization. It is this scale-based adaptability that makes CWT special and analytical in a broad sense.

As an engineering application, the CWT is now an essential tool in any application that involves the detection and characterization of non-stationary processes, including vibration analysis, structural health monitoring, aeronautical engineering, power system fault detection, biomedical signal processing, and interpreting geophysical data [12]. CWT can be used to identify the presence of a bearing defect, gear tooth break, or shaft misalignment by accurately tracking transient perturbations simultaneously in both time and frequency components, which is not possible with Fourier-based methods. Likewise, in aeronautical engineering, where signals due to rotating machinery, airflow disturbances, or structural oscillations rapidly change their spectral characteristics, CWT allows condition monitoring and detection of damage with better resolution [13]. The superiority of CWT over other signal processing techniques (e.g., FT, STFT, Hilbert-Huang Transform) lies in its capacity to support highly localized excitation of events without losing information of global frequencies, making it extremely effective in the context of practical diagnostics. Its adaptive resolution eliminates the resolution trade-off of STFT, and its mathematical representation is redundant but highly interpretable, particularly when feature extraction and fault classification are required [14].

Table 1. Wavelet selection

Mother Wavelet	CWRU	MFS
Haar	0.0768	0.0289
Sym2	0.1011	0.0384
Coif2	**0.1347**	0.0471
More	0.1044	**0.0584**
Db2	0.1011	0.0384

Table 1 indicates a maximum relative energy (MRE) wavelet selection criteria used on CWRU and MFS-bearing data. Maximum Relative Energy (MRE), which may be utilized in the choice of the most suitable mother wavelet to be selected in the decomposition of signals, primarily when non-stationary signals, such as bearing fault diagnosis, signal analysis of gear vibration, are being considered. The principle of MRE is that when a wavelet is modified to an optimal one, the maximum percentage of energy existing in the original signal is retained, so that considerable information about a fault is not lost. Instead of picking the wavelets subjectively or using trial-and-error, MRE gives an evaluation of the performance of the wavelets by comparing the amount of energy picked up by each candidate wavelet (e.g., Daubechies, Symlet, Coiflet, etc.). The method compares the energy of the wavelet coefficients in one level of decomposition and the overall energy of the original signal, with the wavelet that has the largest energy ratio being chosen as the optimum one. The more MRE implies the better time-frequency localization, the better signal-noise ratio, and low information loss that further enhances the extractions of the feature, the more the fault can be separated, and the better the machine learning classification $E_x = \sum_{n=1}^{N} x(n)^2$ and $W_{i,j}$ are the wavelet coefficients with energy $E_{w,j} = \sum_k W_{j,k}^2$. Then the relative energy at level j is $RE_j = \frac{E_{w,j}}{E_x}$ and maximum value among all levels $MRE = max_j \left(\frac{E_{w,j}}{E_x}\right)$. Increased MRE implies an enhanced time-frequency localization, a better signal-to-noise ratio, and less information loss, which subsequently leads to improved feature extraction, fault separability, and classification by machines. Five mother wavelets are considered: Haar, Symlet, Coiflet, Morelet, and Daubechies. As can be seen in Table 1, the largest calculated MRE is 0.1347 (CWRU), the Coiflet wavelet. Conversely, the highest value of MRE obtained represents 0.0584 (MFS) after the application of the Morlet wavelet.

2.2 Feature Selection and Ranking using Fisher Score

Fisher Score, in other words, feature selection and ranking is a powerful statistical method used extensively in machine learning and fault diagnosis because it is simple, robust, and highly discriminative. The approach assesses each feature separately by calculating the ratio of inter-class separation to intra-class variance, so that those features which most effectively discriminate between classes are given preference. The score for the i^{th} feature is computed as:

$$F_i = \frac{\sum_{c=1}^{C} n_c \left(\mu_{i,c} - \mu_i\right)^2}{\sum_{c=1}^{C} n_c \, \sigma_{i,c}{}^2} \tag{1}$$

where $\mu_{i,c}$ and $\sigma_{i,c}{}^2$ are class-wise feature mean and variance, and n_c is the number of samples in class c. A higher Fisher score means better class separability. The most important benefit of the Fisher Score is that it is computationally efficient and does not require any particular classifiers, so it can be applied to high-dimensional data, like vibration signals in bearing fault diagnosis. In contrast to wrapper methods that entail training or computationally intensive embedded methods, Fisher Score is a filter-based method that does not suffer from overfitting and, overall, consumes less processing

time without sacrificing feature relevance [15]. Its principle of operation is to prioritize features based on their ability to discriminate between fault categories. Features with a large Fisher score have high class separability, whereas those with low scores are considered noise or redundant information. This approach is of great value for bearing fault diagnosis, as a vibration signal is usually mixed-frequency and may include redundant features extracted in both the time and frequency domains. The Fisher Score is used to isolate the most sensitive features that represent fault patterns, such as inner race, outer race, or rolling-element defects, which are then used to improve the accuracy of other classifiers, such as SVM, Logistic Regression, or Neural Networks [16]. What makes it unique among feature-ranking methods like ReliefF or mutual information is that it does not need either neighborhood search or probabilistic density estimation, and is thus not sensitive to parameter tuning or changes in sample density. Its ability to effectively and non-critically reduce feature dimensionality in relation to diagnostic performance simplifies real-time implementations, enables generalized models across different operating conditions, and increases overall reliability.

Scalogram images can be analyzed using variance-based features, such as Sum Variance and Sum of Squares Variance, which provide an initial estimate of the spread and distribution of pixel intensities across the time-frequency domain. These characteristics directly point to the unevenness of vibration energy distribution of bearing faults, healthy and faulty conditions having narrower and broader distributions, respectively, with respect to irregular impulses. Continuing this discussion, cluster-based descriptors that extend the above analysis, such as Cluster Shade and Cluster Prominence, measure asymmetry and peak sharpness of scalogram distributions [17]. When bearings are healthy in operation, their textures are usually balanced, but when defects begin to appear, they develop skewed and peaked signatures that are indicative of robust nonlinear modulations as well as repetitive impact signatures. These characteristics are especially important when it is sought to notice early stages of spalling or crack propagation, in which local distortions initially emerge in the time-frequency representation. Adding to these, local variation measures like Difference Variance, Contrast, and Dissimilarity model differences in pixel-to-pixel intensity, and provide sensitivity to abrupt bursts of energy and local fault-based variation. Normal bearings produce less pronounced, more regular textures, though defect development produces more contrasts and anomalies, which are readily appreciable under these measures. Furthermore, to offer a good temporal view, Autocorrelation has the ability to measure the repetitive occurrence of image patterns, which, in essence, help to indicate repetitive fault impulses related to localized outer race, inner race, or rolling element defects. These two characteristics are not only used to determine the existence of faults but also to distinguish between gradual degradation and catastrophic failure [18].

The other important set of characteristics is associated with smoothness and measures of energy distribution. Maximum Probability and Energy express the prevalence of particular textures and the consistency of pixel configurations, both of which are greatly diminished in non-defective circumstances when the energy is scattered throughout the scalogram. The Inverse Difference, Homogeneity, and Inverse Difference Normalized are used to measure the smoothness of texture, that is, how uniform the levels of intensity are distributed. These values are large in a healthy bearing, indicating homogeneity of time-frequency energy, but they are associated with an irregularly fault-induced burst action. Correlation also adds some depth to this view by measuring the extent of linear dependence of pixel intensities. Strong correlation suggests repetitive, regular signals, whereas randomness on the fault causes the relationship to break [19]. Lastly, entropy and information-theoretic measures are even more insightful at fault severity and complexity. Entropy, Sum Entropy, and Difference Entropy are direct measures of the randomness and the disorder of the scalogram and are used as direct measures of instability. Although global entropy indicates general complexity, difference entropy is very sensitive to small irregularities in the early stages of evolution that are not easily noticed using other measures. Inverse Difference Moment Normalized is used to deal with differences in scale in the homogeneity assessment of various operating conditions. Besides, nonlinear and higher-order dependencies are captured in the Information Measure of Correlation 1 and 2, which provide better discrimination in complex situations with overlapping fault types. Being a universal standard, the Mean (Average) intensity is the total energy concentration of the scalogram images, which can easily be compared between normal and abnormal conditions. These variance-based, cluster-based, correlation-based, entropy-based, and smoothness-based features, as a unified set, are effective in detecting and classifying faults reliably in bearing diagnostics using machine learning models [20].

2.3 Machine Learning Methods

Fault diagnosis is an area where Machine Learning (ML) is transforming the world, as it allows automatic feature extraction, adaptive classification, and accurate prediction even in a complicated operating environment. Signals used by bearings are not usually stationary, are full of noise, and are influenced by load, speed, and temperature changes. Typically, the diagnostic pipeline includes signal acquisition, preprocessing, feature extraction (time, frequency, or time-frequency domain), dimensionality reduction, and supervised classification.

2.3.1 Logistic Regression

Logistic Regression is a probabilistic linear classifier that models the conditional probability of a class label using the sigmoid (logistic) function.

$$P(y = 1/x) = \frac{1}{1 + e^{(-\omega^T - b)}} \tag{2}$$

where ω and $\boldsymbol{b}$ are optimized by minimizing cross-entropy loss. LR's key strength is that the interpretability feature weights indicate how input parameters contribute to fault likelihood. LR works best when class boundaries are approximately linear, and datasets are not extremely high-dimensional. While it may not match nonlinear classifiers in accuracy, its probabilistic output supports threshold-based health monitoring and online decision-making in resource-limited industrial systems [21].

2.3.2 Support vector machine

Support Vector Machine, which is a supervised machine learning model that categorizes the data by showing an optimum separation between different classes with a maximal margin. Its mode of operation is to identify the most important data points, colloquially referred to as support vectors, which define this decision boundary. In the case where CWT is applied to transform a signal into time-frequency images or feature vectors, SVM occurs on the resultant features to identify fault types. SVM uses support vectors, or the nearest data points to the decision boundary, in order to identify the best hyperplane to separate two classes. Stability to generalization and lack of overfitting. Margin maximization is primarily used to address small sample problems, which tend to overfit. The decision boundary for a linear SVM is defined as $\boldsymbol{w^T x + b = 0}$, where $\boldsymbol{w}$ is a weight vector, and $\boldsymbol{b}$ is a bias. The optimization problem is written as:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \; subjected\ to\ y_i(w^T x_i + b) \geq 1 \tag{3}$$

For nonlinear problems, SVM uses kernel functions such as RBF, polynomial, or sigmoid:

$$k(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \tag{4}$$

where φ maps data into a higher-dimensional space. With CWT scalograms, SVM operates either on statistical features or deep features.SVM is capable of working on nonlinear complex data in contrast to simple linear classifiers, where data is projected into higher dimensions through the use of a kernel function. This capability enables SVM to learn complex associations among vibration signal characteristics of bearing fault data [22-24]. It is very resilient to noisy measurements and small samples, and is therefore well-suited to industrial diagnostics with limited clean labeled information. SVM is repeatable and gives a consistent result, and does not overfit because it uses only important boundary points and not all data samples. Under bearing fault diagnosis, SVM is applied to the classification of various defects like inner-race, outer-race, and rolling element faults. Even with smaller feature sets to support real-time applications, researchers have reported diagnostic accuracies greater than 98%. The feature extraction methods, such as wavelet transforms, Variational mode decomposition (VMD), and empirical mode decomposition (EMD), have been combined with SVM. This combination can improve its localized damage pattern and eventual vibration detection [25]. It is also doing a good job in detecting faults in the early period, so the maintenance staff can avoid disastrous machine breakages. SVM also has multi-class extensions to deal with the complex fault classification problem with multiple fault classes. Its ability to generalize also makes it so that when trained, it can properly classify previously unobserved operating states. SVM is computationally more expensive than simpler models. However, better solvers and hardware acceleration have made it practical to use in online monitoring systems [26]

2.3.3 K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a very basic yet powerful supervised machine learning algorithm, and it is used in the solution of classification problems. It is founded on the comparison of the new sample of data with those that are kept, and gives the name of the class to the nearest data points to it. In the case of CWT, the features are based on time-frequency coefficients or a salogram, which are input to KNN, and the classification is determined using the similarity of features instead of training models. KNN is based on the fact that similar data points (with regard to their classes) are seen clustering in feature space. The Euclidean distance is normally used to calculate the distance between samples:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{5}$$

KNN does not build a model but stores training data, making it suitable for CWT feature classification where patterns are visually separable. The choice of k influences performance: small k causes noise sensitivity, while large k reduces class separation. KNN is an intuitive pattern recognition technique because the distance measures are used to determine the proximity. Due to its non-parametric nature, KNN does not make any preconceived assumption about the data distribution, and thus can be applied to real-world signals [27]. KNN uses vibration-based sensors in the analysis of defects during bearing fault diagnosis. It is able to clearly distinguish between normal operation and inner-race, outer-race, or rolling element faults. Its main advantage is that it does not need a training phase, and it can be deployed quickly with very low computational costs. KNN can adjust to new data with ease and thus is applicable in fault monitoring systems over the Internet. It is stable where the feature extractors, such as wavelet transform or VMD, are used to sharpen the signal. Since it is making decisions based on real data points, it is inherently interpretable and transparent [28]. KNN is suitable if the dataset size is small or medium-sized and the fault classes are fairly separated in feature space. It can be best used to prove the existence of diagnostic information in selected features.

2.3.4 Bagged tree

Bagged Trees is an ensemble learning technique that enhances the classification accuracy through the combination of the outputs of many decision trees. Their method is based on training large numbers of individual trees on random subsets of training data, known as bootstrap sampling. All trees make their own classification decision, and the overall output is a majority vote of all the trees. At the implementation of CWT to create features, bagged trees are trained to learn many decision boundaries across large sets of features, resulting in fewer overfits and lower variance. Sampling with replacement is the primary principle, according to which several training sets are constructed and each of the trees is trained on its own. By majority voting within trees, the final tree prediction is achieved. If $\boldsymbol{T_1, T_2 \ldots T_m}$ are individual tree predictions, the final bagged classifier is:

$$f(x) = mode\{T_1(x), T_2(x), \ldots T_m(x)\} \quad (6)$$

This method makes the model less susceptible to changes in the data and allows for the reduction of overfitting relative to a single tree. Bagged Trees are used in bearing fault diagnosis to distinguish between the vibration characteristics of a healthy and defective condition [29]. The ensemble effect provides predictive stability even in the case of individual decision trees, making erroneous decisions based on noisy or incomplete signals. They are especially useful when one has to deal with large and heterogeneous datasets, when different operating conditions have to be taken into account. One uncommon benefit is that Bagged Trees inherently give a score of the importance of features that assist engineers in determining the signal parameters that have the biggest impact on fault detection [30]. They do not need significant parameter tuning and can be applied in industrial practice. The reason why Bagged Trees can be applied within online monitoring systems is that they can be quickly trained. The technique has strong resistance to noisy signals, and therefore can be used in rugged conditions where vibration measurements are fragile. It works effectively in combination with feature extraction techniques like wavelet transforms, VMD, or statistical analysis [31]. Since the ensemble aggregates large numbers of weak learners into a single strong learner, the diagnostic accuracy and the diagnostic reliability are always enhanced. Generally, Bagged Trees are appreciated because they are stable, interpretable, and can easily extrapolate to unknown bearing fault conditions.

2.4 Experimental Setup

2.4.1 CWRU data

One of the best-known benchmark datasets currently used to diagnose intelligent faults in rotating machines is the Case Western Reserve University bearing dataset, which provides high-quality vibration data and is under controlled experimental conditions. A specially constructed test rig, comprising a 2-hp Electric motor, a torque transducer, a dynamometer, and a series of deep-groove ball bearings (SKF 6205-2RS JEM), was used to record the data, as shown in Figure 2. The detailed operating parameters and bearing characteristics associated with this experimental setup are summarized in Table 2. In order to model the bearing faults in the real world, electrical discharge machining (EDM) was used to add controlled single-point defects to the inner race, outer race, and rolling elements of the diameters 0.007, 0.014, and 0.021 inches. High-resolution vibration signals with a bandwidth of up to 10 kHz were measured using accelerometers mounted on the drive-end (DE) and fan-end (FE) bearing housings of the motor. The DE and FE represent two distinct sides of the motor where measurements were recorded. As the vibration response varies with bearing position, the present analysis and calculations are focused on the data acquired from the drive-end (DE) bearing. The sampling frequency used was 12 kHz to 48 kHz to ensure that bearing fault signals could be properly captured at different motor speeds of 1797, 1772, 1750, and 1730 rpm. The test rig could be operated under various load conditions between 0 and 3 hp, thus collecting data under various working conditions [32].

Table 2. CWRU Bearing specifications

Sr.no	Bearing Parameters	CWRU
1	Class of bearing	6205 (SKF)
2	Outer race Diameter (mm)	51.99
3	Inner race Diameter (mm)	25.01
4	Size of the ball (mm)	7.94
5	Quantity of the ball	9

The principle of operation was to operate the motor with the given load and the given speed conditions and to adjust the torque with the dynamometer to sustain the same operating conditions. Localized defects are represented by the vibration signals measured at both bearing locations and consist of periodic impulses, which are modulated by shaft speed and bearing geometry. These cues are a perfect source of feature extraction in time, frequency, and time-frequency space. CWRU data is highly useful in bearing fault diagnosis as it provides a controlled, repeatable, and labeled environment, which is suitable to create and test machine learning models, signal processing methods such as VMD, STFT, or CWT, and intelligent diagnostic methods that are a combination of the former. Its exhaustive classification of types of fault, severities, and operating conditions allows researchers to compare their procedures with those of previous researchers in a consistent manner. Additionally, due to the availability of both drive-end and fan-end measurements, the dataset can be used to test the propagation of fault signatures through the shaft system, which is essential to the development of a robust

diagnostic model. Specification--rotor speed, fault size, radial load, and much more detailed--improves the credibility and scientific usefulness of the dataset, and compatibility with MATLAB and Python toolboxes makes signal preprocessing and the recreation of the experiment simpler. Such a combination of exact instrumentation, thorough labeling, and its popularity has made CWRU data a worldwide standard to evaluate the accuracy, reliability, and generalization of bearing fault diagnostic methods in academic and industrial studies.

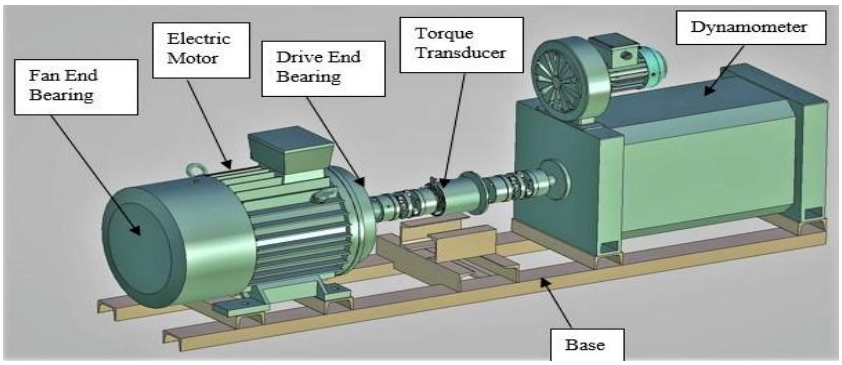

Figure 2. CWRU Bearing test rig (authors' own illustration). [33]

2.4.2 Machinery fault simulator rig

The Machinery Fault Simulator (MFS) rig is a flexible experimental platform that can be used to investigate the dynamics of rotor bearings and obtain vibration measurements under controlled fault conditions. A horizontal rotor was supported with a SKF 6004 deep groove ball bearing in this study, and fault conditions like inner race defect, ball defect, and outer race defect were induced artificially with detailed specifications summarized in Table 3. This configuration also included a DC motor that could serve as a servo to change the shaft speed between 200 rpm and 3000 rpm at 200 rpm intervals, so that systematic behavior analysis could be performed in relation to speed. To record vibration signals in high fidelity, a multi-axis piezoelectric accelerometer of sensitivity 10.33 mV/g was attached to the bearing housing. Sampling frequency was 12 kHz to obtain data that accurately represents defect-induced high-frequency components. The MFS rig is designed with high-accuracy alignment functions to minimize undesired mechanical noise, and the strong shaft-bearing design resembles the actual industry environment. The modular design of the rig allows for replacing test bearings quickly and introducing local defects in order to recreate the experiment. The servo control of the DC motor permits the fine control of the speed, which is necessary to test the dynamic responses in the various operating regimes. The testing of bearings was conducted under normal and faulty conditions to get fully developed datasets. The obtained signals, which include the information regarding structural resonances and defect frequencies, are useful input in the development of machine learning and signal processing algorithms.

Table 3. MFS Bearing specifications

Bearing type	Deep groove ball bearing (MFS)
Outer race diameter (mm)	47
Inner race diameter (mm)	29.31
Size of the ball	7.874
Size of the bearing pitch	33.477
Total number of rollers	8
Contact angle	0
Defect Size (mm)	0.5

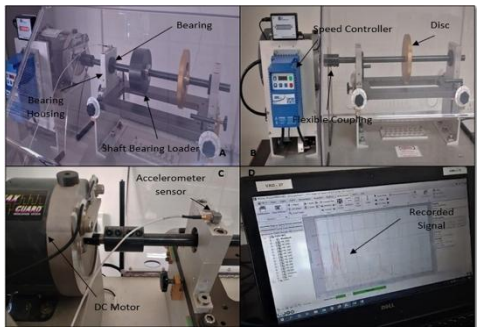

Figure 3. MFS rig and data acquisition

To reduce the effect of external vibrations, the rotor shaft is usually held by a rigid frame, and to ensure the smooth transmission of the torque, coupling mechanisms are provided. High-quality SKF bearing use provides consistency of

results, and the controlled size of defects allows validation of diagnostic models. This methodological approach helps to isolate fault signatures without the interference of other irrelevant sources. The positioning of the accelerometer allows recording radial and axial vibrations, which makes the data set compatible with advanced time-frequency and feature-extraction methods. The sampling rate of 12 kHz is higher than the Nyquist frequency of bearing fault components, so the sampling rate avoids aliasing and allows waveforms to remain intact. The pattern of degradation reported is actual and serves as a reference point to test prognostic and diagnostic tools. The influence of the rotational dynamics on the fault propagation can be examined in detail since the speed can be changed gradually. These holistic data are essential to the design of effective predictive maintenance systems, fault classification programs, and reliability enhancement programs. The MFS rig has proven itself as a dependable platform to conduct research in condition monitoring and bearing fault diagnosis by incorporating high-quality sensors, the control of precise speed, and consistent defect modeling. As shown in Figure 3, the test rig is equipped with a bearing loader, a rotating disc, a flexible coupling, and an accelerometer. The speed controller regulates speed variations between 200 and 3000 rpm. Acceleration is converted into digital signals using a signal conversion device. Furthermore, the signals are saved in the computer for later analysis.

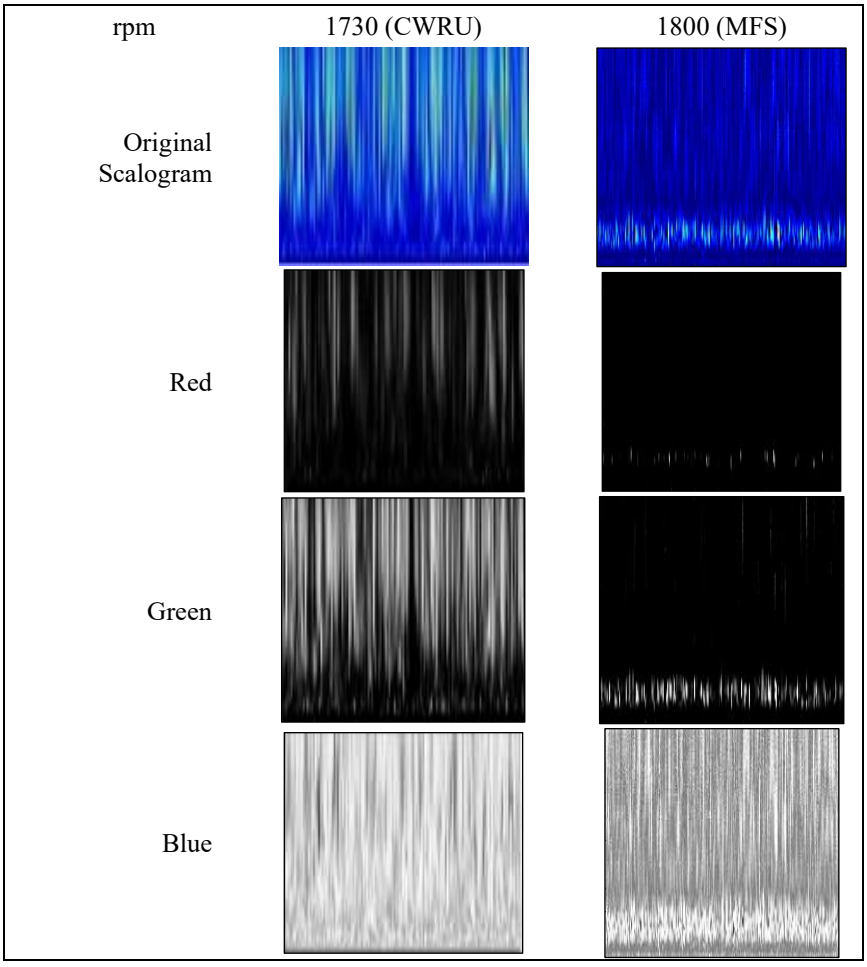

Figure 4. Healthy bearing scalograms

3. Results and Discussion

In this study, the CWRU-bearing data and the MFS-bearing data were used as sources for automatically extracted scalogram information. The 64 vibration signals of CWRU denoting different fault conditions of bearing were transformed into RGB scalograms using the Coiflet wavelet coefficients. The variation of the strength of the RGB scalogram under various fault conditions is observed. Under MFS, 60 vibration signals are recorded in the HB, IRD, ORD, and BD fault conditions. The Morlet wavelet coefficients were converted into scalograms, and all scalograms were finally converted into 180 RGB images. Each RGB scalogram generated several texture features, which were ranked using the Fisher score. Figures 4 through 7 demonstrate the original scalogram and the corresponding RGB scalogram under the same fault conditions at 1730 rpm and 1800 rpm on CWRU and MFS, respectively. The comparison between the CWRU and MFS datasets should be conducted at the same or nearest operating speed to ensure consistency. At this speed, visual differences can be effectively identified. It is observed that pattern variations become more pronounced under different fault conditions. Additionally, the spike amplitudes in the CWRU scalograms are prominently higher compared to those of the healthy bearing scalograms from the MFS dataset. Figure 5. The additional blue-filtered scalogram of CWRU shows greater intensity and color variation than MFS. In the same way, the intensity change of green and red filtered scalograms of CWRU is greater than that of MFS.

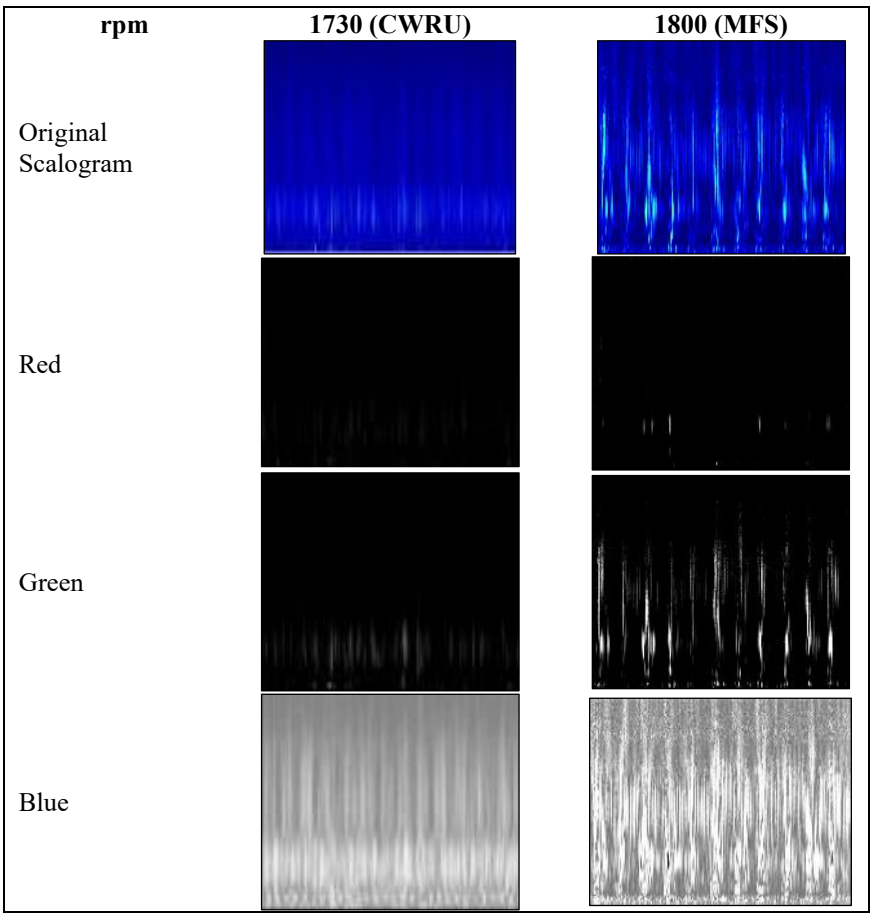

Figure 5. Ball defect scalograms

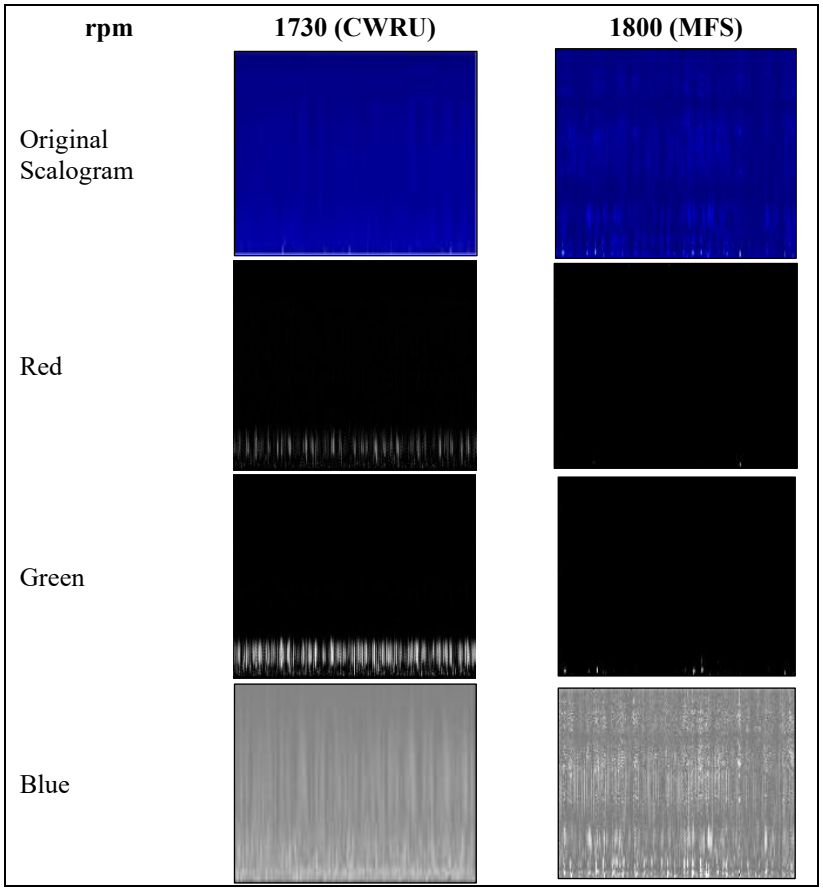

Figure 6. Inner race defect scalograms

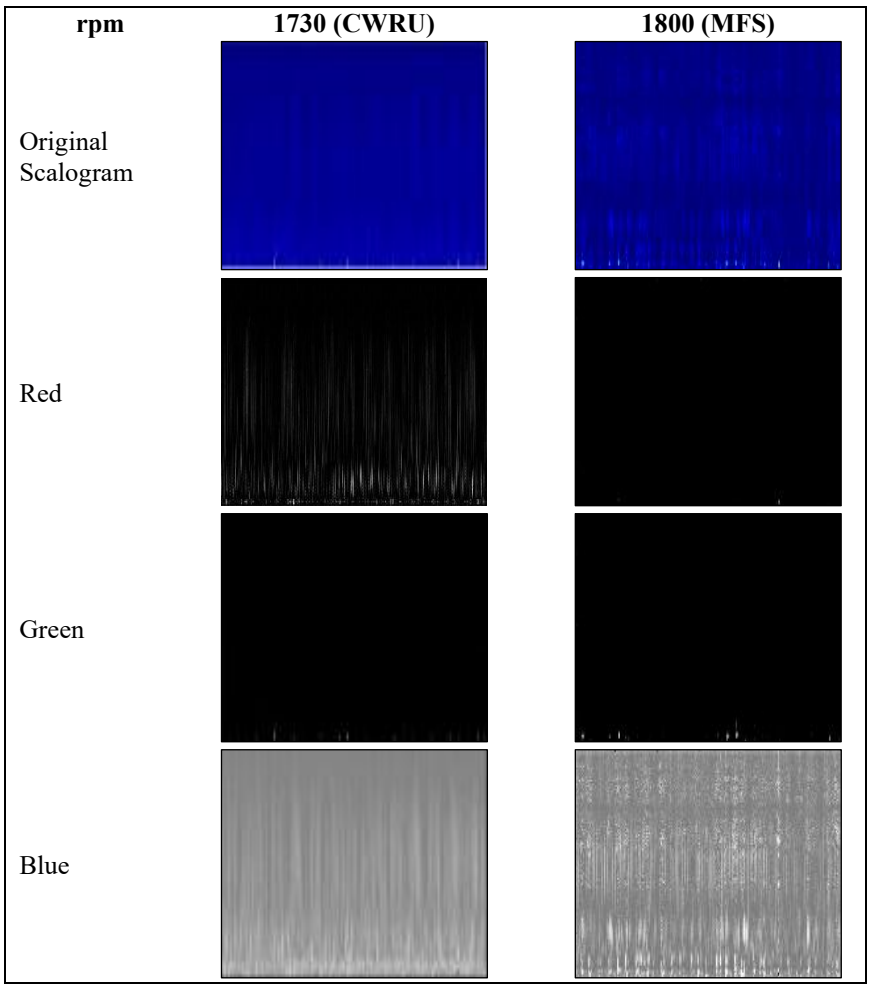

Figure 7. Outer race defect scalograms

The variation in brightness between the BD and MFS scalograms is larger than in CWRU, as in Figure 5. The variation in measurement points among scalograms in IRD and ORD conditions is presented in Figures 6 and 7, respectively. It is noted that, based on the IRD and ORD, greater variability in the red scalogram was noted. IRD condition of CWRU in the green scalogram exhibits a more extreme variation as compared to ORD. Lastly, the blue scalogram of both conditions shows the differences in their images.

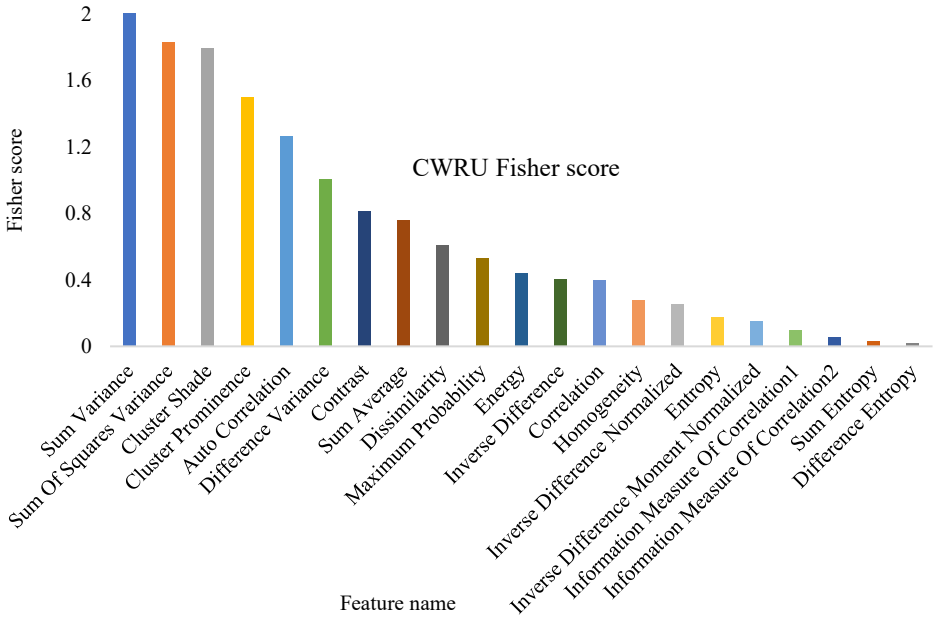

Figure 8. Fisher score ranking for CWRU

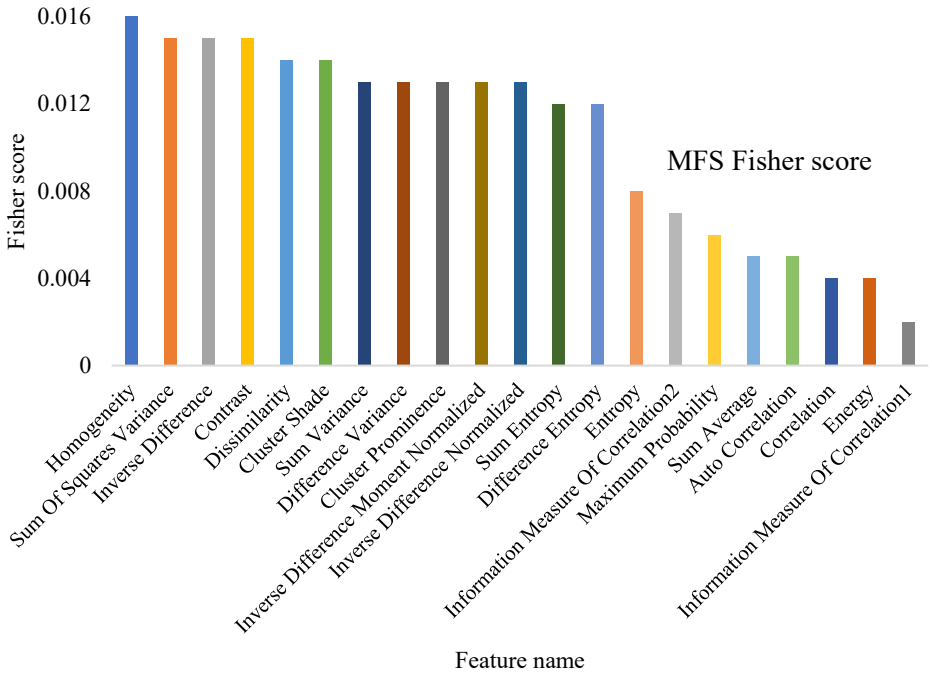

Figure 9. Fisher score ranking for MFS

It has also been noted that images do not provide relevant information for different fault conditions. Thus, the authors obtained 21 texture features of all the RGB scalograms of CWRU and MFS data. Also, Fisher score-based feature ranking has been performed to find the relevant features, and the ranked features are denoted in Figure 8 in the CWRU data. The Feature ranking based on the Fisher score of MFS data is depicted in Figure 9. The feature vectors and training are subjected to classifiers such as SVM, KNN, etc., and the tenfold cross-validation is conducted. The attributes are split into ten matching sets, and ten reiterations are performed on the tenfold cross-validation scheme. Training and tenfold procedures were used to obtain confusion matrices with the considered classification algorithms. The comparison outcome of logistic regression, SVM, KNN, and Bagged Tree as counted on tenfold accuracy at eight and twenty-one features of CWRU and seventeen and twenty-one features of MFS is provided in Table 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{7}$$

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives. This equation is consistent with the confusion-matrix results shown in Table 5, ensuring transparent performance computation across classifiers.

Table 4. Features accuracy

Classifier	Bearing condition	At 8 features	At 21 features	At 17 features	At 21 features
		CWRU		MFS	
Logistic Regression	HB_BD	96.7	88.3	95.6	93.3
	HB_IRD	58.9	91.1	90.0	91.1
	HB_ORD	96.9	94.8	85.6	78.9
SVM	HB_BD	98.3	96.7	83.3	84.4
	HB_IRD	51.1	50.0	83.3	83.3
	HB_ORD	93.8	94.8	76.7	82.2
KNN	HB_BD	98.3	98.3	63.3	84.4
	HB_IRD	85.6	86.7	85.6	88.9
	HB_ORD	97.9	96.9	75.6	84.4
Bagged Tree	HB_BD	98.3	100	77.8	82.2
	HB_IRD	85.6	86.7	87.8	85.6
	HB_BD	98.3	98.3	63.3	84.4

Table 5. Confusion matrix

	at 8th Features	at 21st Features	at 17th Features	at 21st Features
	Confusion matrix (CWRU)		Confusion matrix (MFS)	
	BD 48 0 HB 2 10 BD HB	BD 45 3 HB 4 8 BD HB	BD 45 0 HB 4 41 BD HB	BD 41 4 HB 2 43 BD HB
LR	HB 11 1 IRD 2 46 HB IRD	HB 41 4 IRD 4 41 HB IRD	HB 41 4 IRD 5 40 HB IRD	HB 41 4 IRD 4 41 HB IRD
	HB 12 0 ORD 3 81 HB ORD	HB 11 1 ORD 4 80 HB ORD	HB 39 6 ORD 7 38 HB ORD	HB 36 9 ORD 10 35 HB ORD
	(a)	(b)	(c)	(d)
	BD 48 0 HB 1 11 BD HB	BD 48 0 HB 1 11 BD HB	BD 37 8 HB 7 38 BD HB	BD 36 9 HB 5 40 BD HB
SVM	HB 31 14 IRD 30 15 HB IRD	HB 27 18 IRD 27 18 HB IRD	HB 39 6 IRD 9 36 HB IRD	HB 39 6 IRD 9 36 HB IRD
	HB 7 5 ORD 1 83 HB ORD	HB 8 4 ORD 1 83 HB ORD	HB 33 12 ORD 9 36 HB ORD	HB 36 9 ORD 7 38 HB ORD
	(e)	(f)	(g)	(h)
KNN	BD 48 0 HB 1 11 BD HB	BD 48 0 HB 1 11 BD HB	BD 24 21 HB 12 33 BD HB	BD 36 9 HB 5 40 BD HB
	HB 39 6 IRD 7 38 HB IRD	HB 40 5 IRD 7 38 HB IRD	HB 42 3 IRD 10 35 HB IRD	HB 42 3 IRD 7 38 HB IRD
	HB 11 1 ORD 1 83 HB ORD	HB 11 1 ORD 2 82 HB ORD	HB 38 7 ORD 15 30 HB ORD	HB 37 8 ORD 6 39 HB ORD
	(i)	(j)	(k)	(l)
Bagged Tree	BD 48 0 HB 1 11 BD HB	BD 48 0 HB 0 12 BD HB	BD 37 8 HB 12 33 BD HB	BD 40 5 HB 10 35 BD HB
	HB 39 6 IRD 7 38 HB IRD	HB 40 5 IRD 7 38 HB IRD	HB 42 3 IRD 8 37 HB IRD	HB 42 3 IRD 10 35 HB IRD
	HB 11 1 ORD 1 83 HB ORD	HB 12 0 ORD 1 83 HB ORD	HB 37 8 ORD 11 34 HB ORD	HB 35 10 ORD 13 32 HB ORD
	(m)	(n)	(o)	(p)

The results of the computations were also reported after a 10-fold cross-validation was applied to each subset of fault conditions. A maximum tenfold accuracy of 96.7 percent was attained using Logistic regression using the eighth feature, and a maximum tenfold accuracy of 94.8 percent using twenty-one features using the CWRU data. In the MFS dataset, the highest tenfold accuracy of 95.6% was achieved at 17 features and 93.3% at 21 features. It is also noted that a maximum tenfold accuracy of 98.3% was obtained with SVM using the eighth feature, and a maximum tenfold accuracy of 98.3% was obtained with SVM using twenty-one features on the CWRU dataset. And in the MFS dataset, a maximum tenfold accuracy of 83.3% was achieved at 17 features, and 84.4% at 2/10 features. It was found that the KNN classifier achieved 98.3% accuracy on the CWRU dataset at the eighth and twenty-first features. And in the case of the MFS dataset, the maximum tenfold accuracy of 85.6% is achieved at 17 features, and 88.9% at 21 features. Among the classifiers of the bagged tree, the highest tenfold accuracy of 98.3% was observed for the eighth feature, and the maximum tenfold accuracy of 100% was observed with twenty-one features in the CWRU dataset. And in the MFS dataset, the highest tenfold accuracy of 87.8% is at 17 features, and 85.6% at 21 features. We also found that the HB_BD (Healthy Bearing data compared to Ball Defective data) condition shows better results than all the other conditions. Moreover, to evaluate the accuracy of all ML algorithms and their performance across classes, when error types are more frequent, a confusion matrix is a useful method for summarizing individual accuracy in a tabular format. The correct prediction is indicated by the diagonal elements in the confusion matrix, and the incorrectly predicted results are indicated by blue color, as indicated in Table 5. Based on Logistic regression, we find that there are a few false positives in the CWRU dataset. It is also noted that all the cases of ball deficiency and normal bearing are unquestionably identified in CWRU through Bagged tree classifiers. Tenfold cross-validation can justify the effectiveness of the proposed methodology on defect spots.

4. Conclusions

In this work, the authors applied the Continuous Wavelet Transform to generate RGB scalograms for various fault conditions from the CWRU and MFS data sets. Several texture features based on GLCM were extracted and ranked using the Fisher score to efficiently identify fault conditions. The result can be summarized as follows:
Maximum relative energy criteria are used for five different mother wavelets in both data sets. Coeflet and Morlet wavelets were selected for the CWRU and MFS datasets, respectively. Four classifiers and three pairs of different bearing-fault conditions are compared using tenfold cross-validation.

(i) The maximum tenfold accuracy was observed in both datasets for the eighth and seventeenth features under HB_BD conditions.

(ii) Maximum accuracy of 98.3% is observed with eight ranked features, and 100% with all 21 features extracted from the CWRU dataset.

(iii) Maximum accuracy of 95.6% is observed with 17 ranked features, and 93.3% with all 21 features extracted from the MFS dataset.

(iv) RGB scalograms generated from the CWRU dataset achieve better tenfold accuracy than those from the MFS dataset across all fault conditions.

(v) The bagged Tree classifier gives a maximum tenfold accuracy of 100 %.

Acknowledgements

The authors would like to thank Kenneth A. Loparo, Department of Electrical, Computer, and Systems Engineering, Case School of Engineering, Case Western Reserve University, for providing access to the dataset for research purposes. Further, the authors would like to acknowledge the PDEU Gandhinagar for providing the test rig facility for the MFS data set.

Funding

This study was not supported by any grants from funding bodies in the public, private, or not-for-profit sectors.

Declaration of Competing Interest

The author declares no conflicts of interest.

CRediT Authorship Contribution Statement

Vipul Dave (Conceptualization; Methodology; Validation; Formal analysis; Investigation; Writing—original draft preparation; Supervision),
Vinay Vakharia (Methodology; Data curation; Formal analysis),
Pradeep Kumar Karsh (Formal analysis; Validation; Resources),
Vipul M. Dabhi (Software; Validation; Resources),
Khyati Zalawadia (Data curation; Resources; Investigation).

Availability of Data and Materials

The data supporting this study's findings are available on request from the corresponding author.

Ethics Declarations

This study did not involve human participants or animals. Ethical approval was therefore not required.

Generative Artificial Intelligence Declarations

The authors stated that generative AI was not used to generate content, ideas, or theories. We have just utilized AI to enhance readability and refine the language. This was used with extreme human control and oversight. The authors take full responsibility for reviewing and approving the content.

References

[1] Z. Zhai, L. Luo, Y. Chen, X. Zhang, "Rolling bearing fault diagnosis based on a synchrosqueezing wavelet transform and a transfer residual convolutional neural network". *Sensors,* 25(2), p.325, 2025.

[2] H. Lv, Z. Dong, X. Li, "A fault diagnosis method for gas turbine rolling bearings with variable speed based on dynamic time-varying response and joint attention mechanism". *Sensors*, 25(21), 6617, 2025.

[3] Y. Li, C.Liu, W. Wang, "A bearing fault diagnosis method based on M-SSCNN and multi-layer attention mechanism". *Structural Health Monitoring*, 24(1), 2025.

[4] P.K. Sahu R.N. Rai, "Fault diagnosis of rolling bearing based on an improved denoising technique using complete ensemble empirical mode decomposition and adaptive thresholding method," *Journal of Vibration Engineering and Technologies,* vol. 11, no. 2, pp. 513–535, 2023.

[5] J. Du, X. Li, Y. Gao, L. Gao, "Integrated Gradient-Based Continuous Wavelet Transform for Bearing Fault Diagnosis," *Sensors*, vol. 22, no. 22, p. 8760, 2022.

[6] S. Li, Y. Peng, Y. Shen, S. Zhao, H. Shao, G. Bin, et al., "Rolling-bearing fault diagnosis under data imbalance and variable speed based on adaptive clustering weighted oversampling," *Reliability Engineering and System Safety,* vol. 242, p. 109010, 2023

[7] Y. Ma, J. Cheng, P. Wang, J. Wang, Y. Yang, "Rotating Machinery Fault Diagnosis Based on Multivariate Multiscale Fuzzy Distribution Entropy and Fisher Score," *Measurement*, vol. 2021, p. 109495, 2021.

[8] J. Zhou, M. Xiao, Y. Niu, G. Ji, "Rolling bearing fault diagnosis based on WGWOA-VMD-SVM," *Sensors*, vol. 22, p. 6281, 2022.

[9] A. Orhan, N. Yordanov, M. Ertargin, M. Zhilevski, M. Mikhov, "A Comparative Study of Time–Frequency Representations for Bearing and Rotating Fault Diagnosis Using Vision Transformer", *Machines*, vol. 13, no. 8, p. 737, 2025.

[10] J. Liang, X, Mao, "Rectifier fault diagnosis based on euclidean norm fusion multi-frequency bands and multi-scale permutation entropy." *Electronics*., vol. 14.3, p. 612, 2025.

[11] Y. Zhang, S. Cheng, J. Luo, "Rolling bearing fault diagnosis using continuous wavelet transform and convolutional neural network," *Mechanical Systems and Signal Processing*, vol. 172, p. 109034, 2022.

[12] V. Suthar, V. Vakharia, V.K. Patel, M. Shah, "Detection of Compound Faults in Ball Bearings Using Multiscale-SinGAN, Heat Transfer Search Optimization, and Extreme Learning Machine," *Machines*, vol. 11, no. 1, p. 29, 2023.

[13] J. Du, X. Li, Y. Gao, L. Gao, "Integrated Gradient-Based Continuous Wavelet Transform for Bearing Fault Diagnosis," *Sensors*, vol. 22, p. 8760, 2022.

[14] B. Kim, J. Lee, "Fault Diagnosis and Noise Robustness Comparison of Rotating Machinery using CWT and CNN," *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 1, pp. 1279-1285, 2021

[15] K. Gao, Z. Wu, C. Yu, "Composite fault diagnosis of rolling bearings: A feature selection approach based on the causal feature network," Appl. Sci., vol. 13, no. 16, p. 9089, 2023.

[16] M. Wang, Z. Zhang, X. Li, "Bearing fault detection by using graph autoencoder and ensemble learning," *Sensors*, vol. 24, no. 21, p. 7589, 2024.

[17] M.F. Siddique, M.A. Ganaie, S.S.S.R. Depuru, "Advanced bearing-fault diagnosis and classification using Mel-transformed scalograms obtained from vibrational signals," *Journal of Vibration Engineering and Technologies*, vol. 12, no. 1, pp. 1931–1942, 2024.

[18] M. Altaf, M. A. Ganaie, and S. S. S. R. Depuru, "A new statistical features based approach for bearing fault diagnosis using vibration signals," *Applied Sciences*, vol. 13, no. 16, p. 9089, 2023.

[19] R.K. Mishra, A. Choudhary, S. Fatima, A.R. Mohanty, B.K. Panigrahi, "A self-adaptive multiple-fault diagnosis system for rolling element bearings." *Measurement Science and Technology*., vol. 33.12, p. 125018, 2022.

[20] J. Yang, C. Zhou, X. Li, "A novel feature extraction method based on symbol-scale entropy for rolling bearing fault diagnosis," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 237, no. 15, pp. 3582–3595, 2023.

[21] Y. Gao, Z. Ahmad, J.-M. Kim, "Fault Diagnosis of Rotating Machinery Using an Optimal Blind Deconvolution Method and Hybrid Invertible Neural Network," *Sensors*, vol. 24, no. 1, p. 256, 2024.

[22] H.P. Raturi, S. Kushari, P.K. Karsh, S. Dey, "Evaluating stochastic fundamental natural frequencies of porous functionally graded material plate with even porosity effect: a multi-machine learning approach," *Journal of Vibration Engineering & Technologies*, vol. 12, no. 2, pp. 1931-1942, 2024.

[23] Vaishali, P.K. Karsh, S. Kushari, R.R. Kumar, S. Dey**,** "Stochastic free vibration and impact responses of functionally graded plates: a support vector machine learning model approach," *Journal of Vibration Engineering and Technologies*, vol. 11, no. 7, pp. 2927–2943, 2023.

[24] P. K. Karsh, S. Dey, “Comparison of multiple surrogate models probing uncertainty in natural frequency of hybrid functionally graded sandwich cylindrical shells,” *Journal of Mechanics of Materials and Structures*, vol. 17, no. 2, pp. 97–121, 2022.
[25] N. Lei, F. Huang, C. Li, "Rolling bearing fault diagnosis based on variational mode decomposition and weighted multidimensional feature entropy fusion." *Journal of Vibroengineering,* vol. 26.3, p. 590-614, 2024.
[26] H. Luo, Y. Zhang, X. Liu, "A lightweight and small sample bearing fault diagnosis algorithm based on probabilistic decoupling knowledge distillation and meta-learning," *Journal of Vibration Engineering and Technologies*, vol. 12, no. 2, pp. 1931–1942, 2024
[27] W. Li, Y. Zhang, Z. Zhang, "An orthogonal wavelet transform-based K-nearest neighbour classifier for bearing fault diagnosis," *Computers and Electrical Engineering*, vol. 98, p. 107585, 2022.
[28] J. Lu, X. Zhang, Y. Zhang, "Enhanced K-nearest neighbour for intelligent fault diagnosis of rolling bearings," *Electronics*, vol. 11, no. 3, p. 919, 2021.
[29] A. Alhams, A. Abdelhadi, Y. Badri, S. Sassi, J. Renno, "Enhanced bearing fault diagnosis through trees ensemble method and feature importance analysis," *Journal of Vibration Engineering and Technologies*, vol. 12, no. 1, pp. 1931–1942, 2024.
[30] Z. Mian, Y. Zhang, H. Zhang, "A literature review of fault diagnosis based on ensemble learning," *Journal of Intelligent Manufacturing,* vol. 35, no. 5, pp. 1245–1263, 2024.
[31] H. Yang, Y. Zhang, X. Li, "Deep ensemble learning with non-equivalent costs of fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 146, p. 107019, 2021.
[32] S. Gao, J. Liu, Z. Zhang, "Fault diagnosis of rolling bearings based on improved energy entropy and vibration amplitude attenuation," *Mechanical Systems and Signal Processing*, vol. 157, p. 107661, 2021.
[33] K.A. Loparo, M.L. Adams, W. Lin, M.F. Abdel-Magied, N. Afshari, "Fault detection and diagnosis of rotating machinery," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1005–1014, 2000.