Leakage-Aware Multi-Condition Fault Diagnosis of Centrifugal Pumps Using Sensor-Sensitive Vibration Features, Random Forest, and Multichannel 1D-CNN

Samed Saeed
VTU Research Resource Center
Kalaburagi, Visvesvaraya Technological University, Belagavi-590018, Karnataka
India

Babu Reddy
VTU Research Resource Center
Kalaburagi, Visvesvaraya Technological University, Belagavi-590018, Karnataka
India

Reliable centrifugal-pump fault diagnosis requires accurate classification supported by physically interpretable evidence. This study presents an integrated vibration-based framework combining analytical signal processing, sensor-sensitivity analysis, trial-level statistical validation, feature-based machine learning, and a five-channel one-dimensional convolutional neural network (1D-CNN). The University of Twente Motor-2 dataset, comprising ten health conditions and three operating levels, was analysed using five accelerometer channels. Following preprocessing and quality control, 948 independent trials were used for paired model development. Statistical analysis showed that 74.4% of fault-versus-Normal comparisons remained significant after Benjamini–Hochberg correction, while 67.4% exhibited medium or large Cliff's delta effects. Across five training repetitions, the 1D-CNN achieved mean window- and trial-level accuracies of ***99.76 ± 0.26%*** *and* ***99.90 ± 0.23%****, respectively. On 191 independent test trials, the speed-excluded Random Forest achieved* ***98.95%*** *trial accuracy, whereas the selected CNN correctly classified all trials. The proposed framework integrates interpretable vibration analysis with robust multichannel deep learning for reliable centrifugal-pump fault diagnosis.*

Keywords: *Centrifugal pump; Condition monitoring; Vibration analysis; Sensor sensitivity; Machine learning; 1D-CNN; Fault diagnosis; Empirical mode decomposition*

1. INTRODUCTION

Centrifugal pumps are essential components in power generation, chemical processing, petroleum production, water distribution, marine systems, and numerous other industrial applications. Their continuous and reliable operation is critical because unexpected failures can interrupt production, increase maintenance costs, reduce energy efficiency, and compromise operational safety. Common faults include bearing defects, impeller damage, rotor abnormalities, looseness, misalignment, blockage, and motor-related degradation. Consequently, condition monitoring and predictive maintenance have become important research areas for improving pump reliability and reducing unplanned downtime. Recent review studies have highlighted the growing adoption of intelligent fault diagnosis techniques that integrate vibration analysis with machine learning for predictive maintenance of pumping systems [1–3].

Among the available monitoring techniques, vibration analysis remains one of the most effective approaches because most mechanical and electromechanical faults directly alter the dynamic response of rotating machinery. Changes in vibration amplitude, impulsiveness, spectral content, modulation characteristics, and time-varying energy distribution provide valuable information about machine health. Rodrigues et al. demonstrated the effectiveness of vibration signals for electric submersible pump fault diagnosis [4], while Varejão et al. emphasized the importance of publicly available benchmark datasets for reproducible evaluation of intelligent diagnostic algorithms [5].

Traditional vibration-based diagnosis relies on analytical signal processing and handcrafted feature extraction. Time-domain descriptors, including root mean square, peak-to-peak value, kurtosis, and crest factor, characterize vibration energy and impulsive behaviour, whereas frequency-domain and transformed-domain features provide complementary information about spectral distribution and non-stationary signal characteristics. Previous studies have shown that appropriate feature extraction and dimensionality reduction can significantly improve centrifugal pump fault discrimination [6–8]. Other investigations have demonstrated that combining vibration signals with alternative sensing modalities, such as motor current, can further enhance diagnostic capability [9]. Similar observations have also been reported for other rotating machinery applications using statistically selected vibration features [10].

Machine-learning algorithms have become increasingly popular because they can learn complex rela–

Received: July 2026, Accepted: September 2026
Correspondence to: Samed Saeed, VTU Research Resource Center, Kalaburagi, Visvesvaraya Technological University, Belagavi-590018, Karnataka, India.
E-mail: samedsaeed3@gmail.com
doi: 10.5937/fme2604685S
© Faculty of Mechanical Engineering, Belgrade. All rights reserved

tionships between engineered vibration features and machine health conditions. Random Forest, support vector machine, and other conventional classifiers have demonstrated reliable diagnostic performance for centrifugal pump monitoring while maintaining relatively low computational cost and good physical interpretability [11–13]. Nevertheless, their performance depends strongly on the quality of manually engineered features and their robustness under varying operating conditions.

Recent advances in deep learning have reduced the dependence on handcrafted feature engineering by enabling automatic extraction of discriminative representations from raw or transformed vibration signals. Convolutional neural networks, convolutional autoencoders, and transfer-learning approaches have shown promising performance for centrifugal pump diagnosis using vibration scalograms, Hilbert–Huang representations, and other time–frequency images [14–20]. Although these methods often report excellent classification accuracy, high predictive performance alone is insufficient unless appropriate validation strategies are adopted and the learned diagnostic decisions remain physically interpretable.

Another important consideration is sensor placement. Because fault-induced vibration propagates through different structural paths before reaching each accelerometer, measurement location strongly influences the observed vibration characteristics. Several studies have demonstrated that sensor position and channel selection significantly affect diagnostic performance and have proposed attention-based or sensor-fusion approaches to exploit complementary measurement information [21–23]. These findings indicate that sensor selection should be treated as an integral part of the diagnostic framework rather than merely an experimental configuration.

Despite substantial progress, recent studies have continued to refine both feature-based diagnosis and model-validation strategies. Dave et al. [24] combined time–frequency vibration analysis, LASSO-based feature selection, and Random Forest classification for bearing-fault identification, demonstrating the continued relevance of interpretable feature-based approaches for rotating-machinery diagnosis. Yang et al. [25] addressed fault diagnosis under unknown operating conditions using a domain-adaptation framework for marine centrifugal pumps. Cabrera et al. [26] employed adversarial one-class learning for fault detection in a multistage centrifugal pump, while Varejão et al. [27] highlighted the similarity-bias problem in vibration-based intelligent fault diagnosis and the importance of maintaining independence between related recordings during model evaluation.

Yin et al. addressed class imbalance in aviation fuel-pump diagnosis through adaptive weighting [28], while Liu et al. proposed a sample-enhancement strategy for rotating machinery under class-imbalanced conditions [29]. Deng et al. further investigated cost-sensitive knowledge transfer for imbalanced fault diagnosis [30]. From an interpretability perspective, Bromberger et al. employed interpretable machine learning and Shapley-based explanations for industrial water-pump failure prediction [31]. Although these approaches improve particular aspects of diagnostic robustness, class imbalance, or interpretability, they generally address these challenges individually rather than integrating statistical fault discrimination, sensor-location sensitivity, and leakage-controlled model evaluation within a common vibration-diagnostic framework.

Other recent studies have focused on advanced diagnostic architectures and deployment-oriented approaches. Xu et al. developed a digital-twin-based centrifugal-pump fault-diagnosis system incorporating transfer learning and graph convolutional neural networks [32]. Jiang et al. addressed multilabel compound-fault diagnosis using a deep-learning framework for centrifugal pumps [33]. Kumar et al. applied an InceptionV3-based approach to blockage-fault diagnosis [34], whereas Dias et al. investigated edge-based intelligent fault diagnosis for centrifugal pumps [35]. Saxena and Rani combined vibration and thermal information for fault detection in polymer pump impellers [36]. More recently, Li et al. investigated automated machine learning using multimodal Mel-spectrogram and vibration information [37]. These studies demonstrate substantial progress in digital twins, compound-fault recognition, image-based deep learning, edge implementation, and multimodal diagnosis; however, quantitative fault-specific sensor sensitivity supported by independent trial-level statistical significance and effect-size analysis is not generally integrated with direct machine-learning and deep-learning comparison.

A particularly relevant recent development was reported by Ataç et al. [38], who analysed five-channel centrifugal-pump vibration measurements using time- and frequency-domain features and multiple machine-learning classifiers with five-fold Group KFold cross-validation based on recording identifiers. Their study represents an important step toward leakage-aware evaluation and also relates fault diagnosis to energy consumption and CO□-equivalent emissions. Nevertheless, the framework remains feature-based and does not integrate grouped validation with trial-level fault-versus-Normal effect-size analysis, fault-specific sensor-feature sensitivity ranking, and direct comparison with a raw multichannel deep-learning model under identical grouped partitions.

Based on these observations, several research gaps remain. First, classification performance is frequently emphasized without complementary statistical verification that vibration characteristics provide both statistically significant and practically meaningful fault-versus-Normal separation. Second, although multi-sensor measurements and sensor-fusion methods are increasingly employed, quantitative fault-specific sensor sensitivity based on independent trials, effect-size magnitude, and operating-level consistency remains less systematically investigated. Third, feature-based machine learning and raw-signal deep learning are commonly evaluated as separate diagnostic approaches, limiting direct comparison under identical independent data partitions. Accordingly, the present study develops an integrated vibration-based fault-diagnosis framework using the publicly available University of Twente Motor-2 centrifugal-pump dataset [39]. The framework combines independent trial-level statistical validation,

fault-specific sensor–feature sensitivity assessment, multi-domain vibration interpretation, and leakage-aware predictive evaluation. An interpretable speed-excluded Random Forest and a raw five-channel 1D-CNN are finally evaluated using identical grouped trial partitions, enabling direct paired comparison while maintaining independence between training and test trials.

The proposed framework is also positioned with respect to recent centrifugal pump diagnostic appro–aches. Ullah et al. demonstrated the usefulness of wavelet-coherence representations and principal component analysis for pump fault classification [40]. Hu et al. combined Boruta-SHAP feature selection with an enhanced stacked denoising autoencoder to improve coupled fault diagnosis and model interpretability [41]. Koç and Saçaklıdır developed an optimized ensemble-learning approach using acoustic signals for centrifugal pump diagnosis [42]. Unlike these studies, the present work integrates sensor-sensitivity evaluation, analytical vibration interpretation, trial-level statistical validation, leakage-aware grouped learning, and paired comparison between conventional machine learning and deep lear–ning within a single diagnostic framework.

The principal contributions of this study are sum–marized as follows:

A trial-level sensor-sensitivity framework is develo–ped to identify fault-specific high-sensitivity accele–rometer locations using statistical significance, Cliff's delta effect size, and operating-level consistency.

Multi-domain analytical vibration analysis is performed to characterize fault-dependent behaviour in the time, frequency, envelope, time–frequency, and em–pirical-mode domains.

Trial-level statistical validation is introduced to verify both statistical significance and practical effect size while avoiding dependence among overlapping vibration windows.

A speed-excluded Random Forest classifier is evalu–ated using grouped trial-level validation to reduce short–cut learning associated with operating-level information.

A multichannel raw-signal 1D-CNN is developed and directly compared with the Random Forest using identical independent test trials and paired statistical evaluation.

The proposed methodology combines the physical interpretability of analytical vibration analysis with the automatic feature-learning capability of deep learning. By integrating sensor-sensitivity analysis, leakage-awa–re validation, statistical effect-size assessment, and paired model comparison, the framework provides a rigorous and practically interpretable approach for cent–rifugal pump fault diagnosis under laboratory operating conditions.

2. METHODOLOGY

2.1 Experimental Dataset and Health-Condition Description

The vibration data used in this study were obtained from the publicly available University of Twente pump condition monitoring dataset [39]. The Motor-2 subset was selected because it contains measurements at three represented operating levels, namely 50%, 75%, and 100%, and therefore allows the influence of operating condition to be considered during fault diagnosis. The dataset was acquired from a motor-driven centrifugal pump system using five accelerometers mounted at different locations on the pump–motor assembly. The considered health conditions were Normal, New motor baseline, Bearing BPFI, Bearing BPFO, Bearing fault, Impeller damage, Loose foot, Soft foot, Rotor fault, and Stator fault.

The five accelerometer locations were as follows: Channel 1, motor non-drive-end bearing in the hori–zontal direction; Channel 2, motor drive-end bearing in the vertical direction; Channel 3, motor drive-end bearing in the axial direction; Channel 4, pump drive-end bearing in the horizontal direction; and Channel 5, pump non-drive-end bearing in the vertical direction. This multi-sensor arrangement enabled evaluation of sensor-location effects on fault detectability.

The original Motor-2 subset contained 380 vibration files, of which 312 were successfully processed. The remaining 68 files were excluded because none of the trials detected within those files satisfied the required common analysis duration used for standardized signal segmentation and feature extraction. The files them–selves were successfully imported and contained detec–table signal columns; therefore, their exclusion reflected the common-duration criterion rather than file corrup–tion or import failure.

After signal-quality screening, 54,342 valid channel-window feature observations were retained for ana–lytical and statistical investigations. The initial comp–lete-channel cohort contained 1,033 independent trial groups and 46,485 complete-channel feature observa–tions. Raw five-channel reconstruction failed for 84 trial groups, and one additional trial group was removed during gross-amplitude quality control. The final com–mon modelling cohort therefore contained 948 inde–pendent trials, 42,660 feature observations for conven–tional machine learning, and 8,532 synchronized five-channel windows for CNN development. The dataset summary and health-condition categories are provided in Tables 1 and 2.

Table 1. Summary of the Motor-2 dataset and final modelling cohort

Parameter	Value
Experimental configuration	Motor-2
Represented operating levels	50%, 75%, and 100%
Number of accelerometers	5
Health conditions	10
Original vibration files	380
Successfully processed files	312
Valid channel-window feature observations	54,342
Initial complete-channel feature observations	46,485
Initial complete five-channel trial groups	1,033
Raw five-channel reconstruction failures	84
Gross-amplitude quality-control exclusions	1
Final common independent trial groups	948
Final paired ML feature observations	42,660
Final synchronized CNN windows	8,532
Approximate source-recording duration	12 s

2.2 Data Preparation and Signal Segmentation

The Motor-2 recordings were organized according to operating level, health-condition label, sensor channel, and experimental trial identifier. A metadata inventory was generated to preserve traceability between each feature observation, source file, sensor channel, operating level, and trial group. Signals with missing measurements, insufficient duration, corrupted values, incomplete channel coverage, or physically unrealistic responses were excluded.

Table 2. Health conditions considered in this study

Health condition	Condition type
Normal	Baseline
New motor baseline	Baseline
Bearing BPFI	Mechanical
Bearing BPFO	Mechanical
Bearing fault	Mechanical
Impeller damage	Hydraulic
Loose foot	Mechanical
Soft foot	Mechanical
Rotor fault	Electrical/Mechanical
Stator fault	Electrical

To prevent information leakage, repeated windows originating from the same experimental trial were assigned to the same trial group. No trial group was allowed to appear in more than one of the training, validation, and testing subsets. Accepted vibration records were standardized to a common analysis duration and segmented using a sliding-window procedure. For a window length (L), overlap ratio (r), and signal length (N), the step size and number of windows were calculated as:

$$S = L(1-r) \tag{1}$$

$$N_w = \left\lfloor \frac{N-L}{S} \right\rfloor + 1 \tag{2}$$

where (N_w) represents the total number of extracted windows and $\lfloor . \rfloor$ denotes the floor function.

In the analytical feature branch, a 2 s window and 50% overlap were used. With an approximately 10 s standardized analysis interval, each trial-channel record produced nine overlapping windows. After feature extraction and quality screening, the final paired machine-learning cohort contained 948 independent trial groups and 42,660 feature observations. For the CNN branch, raw five-channel signals were reconstructed over a common 5 s interval and segmented into 1 s synchronized windows with 50% overlap, producing nine windows per trial.

The complete preprocessing sequence is illustrated in Figure 2. Raw multichannel vibration signals were synchronized and resampled before being restricted to a common analysis duration and segmented into overlapping windows. Analytical features were extracted from individual channel windows. Trial-level medians were subsequently computed for statistical validation and sensor-sensitivity assessment, whereas the window-level channel-feature observations were retained for conventional machine-learning analysis.

Table 3. Summary of preprocessing outcomes

Processing stage	Number
Original Motor-2 files	380
Successfully processed files	312
Initially generated channel-window feature observations	54,513
Observations removed during anomaly screening	171
Valid channel-window feature observations	54,342
Initial complete five-channel trial groups	1,033
Initial complete-channel feature observations	46,485
Valid observations excluded because of incomplete channel coverage	7,857
Raw five-channel reconstruction failures	84 trials
Gross-amplitude quality-control exclusions	1 trial
Final common independent trial groups	948
Final paired ML feature observations	42,660
Final synchronized five-channel CNN windows	8,532
Training / validation / testing trials	644 / 113 / 191

2.3 Analytical Signal Analysis

Analytical signal processing was used to interpret the physical vibration behaviour of the different health conditions and to generate features for statistical validation and conventional machine-learning classification. The extracted features represented complementary vibration characteristics in the time, frequency, envelope, time–frequency, and empirical-mode domains.

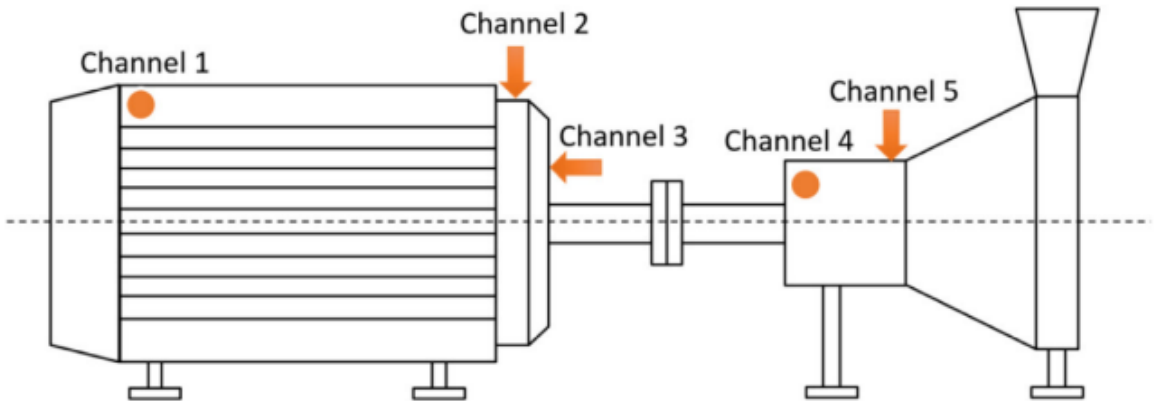


Figure 1. Schematic representation of the Motor-2 pump–motor experimental setup showing the five accelerometer locations used for vibration acquisition [39].

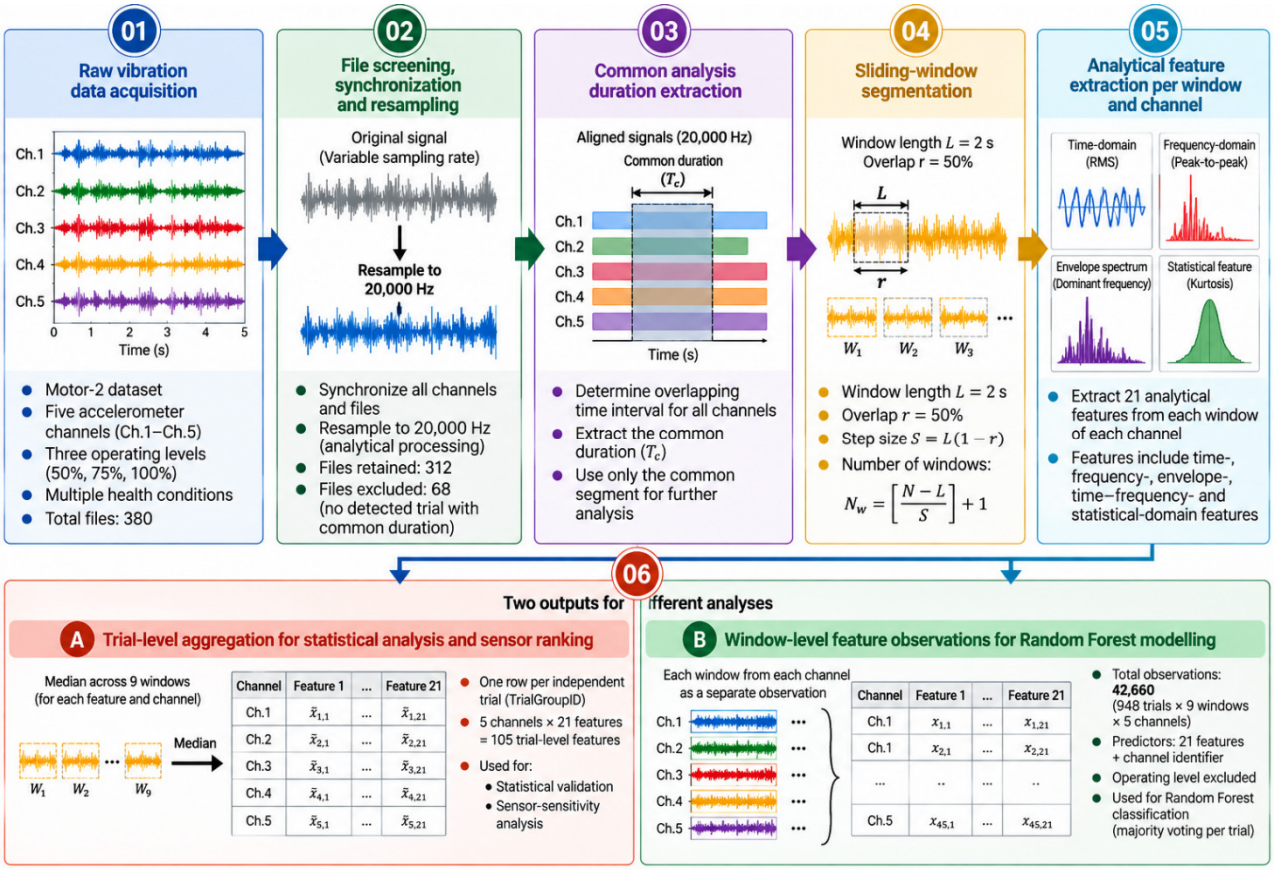


Figure 2. Workflow for vibration-signal preparation, analytical feature extraction, trial-level statistical aggregation, and window-level Random Forest modelling.

Time-domain analysis included root mean square, peak-to-peak amplitude, kurtosis, and crest factor to characterize vibration energy and impulsiveness. Frequency-domain analysis was performed using the fast Fourier transform and Welch power spectral density estimation to examine spectral-energy distribution. Envelope analysis was carried out using the Hilbert transform to enhance amplitude-modulated and impact-related components. Time–frequency behaviour was analysed using the short-time Fourier transform, and empirical mode decomposition was applied to separate the vibration signal into intrinsic mode functions.

For a discrete vibration signal x(n), the principal feature definitions were retained as follows:

$$RMS = \sqrt{\left(\frac{1}{N}\right)\sum_{n=1}^{N} x_n^2} \tag{3}$$

$$x_{p2p} = \max\left(x_n\right) - \min\left(x_n\right) \tag{4}$$

$$CF = \frac{\max\left(\left|x_n\right|\right)}{RMS} \tag{5}$$

The empirical mode decomposition of a signal x(t) was expressed as

$$x(t) = \sum_{t=1}^{M} c_i(t) + r(t) \tag{6}$$

where $c_i(t)$ is the i^{th} IMF and r(t) is the residual. The normalized energy of the i^{th} IMF was calculated as

$$E_i^{\text{norm}} = \frac{\sum_{n=1}^{N} c_i^2(n)}{\sum_{j=1}^{m}\sum_{n=1}^{N} c_j^2(n)} \tag{7}$$

The objective of the analytical stage was not only classification but also interpretation of fault-induced vibration behaviour. The analytical techniques and their diagnostic roles are summarized in Table 4.

Table 4. Summary of the analytical techniques and their objectives

Method	Purpose	Output
Time-domain analysis	Evaluate signal amplitude and impulsiveness	RMS, Kurtosis, Crest Factor
Frequency-domain analysis	Identify spectral signatures	FFT, PSD
Envelope analysis	Detect repetitive impacts	Envelope spectrum
STFT	Analyse non-stationary behaviour	Spectrogram
EMD	Decompose nonlinear signals	IMFs, IMF energy
Trial-level statistical validation	Assess distributional differences among health conditions using independent trial medians with FDR correction	Kruskal–Wallis (H), adjusted p-values
Pairwise effect-size analysis	Compare each condition with Normal at the same operating level and channel	FDR-adjusted rank-sum (p), signed Cliff's δ

2.4 Trial-Level Statistical Validation and Sensor-Sensitivity Analysis

Because the nine overlapping windows extracted from a single trial are not statistically independent, formal statistical inference was performed after trial-level aggregation. For each trial, sensor channel, operating level, and analytical feature, the nine window-level feature values were summarized using the median:

$$\tilde{x}_{i,c,m,s} = median_{w=1,\ldots,9}\left(x_{i,c,m,s,w}\right) \quad (8)$$

where $\tilde{x}_{i,c,m,s}$ denotes the value of feature m obtained from window w of trial i, recorded using channel c at operating level s. The final statistical cohort contained 948 independent trials and five sensor channels, producing 4,740 trial-channel summaries.

Statistical testing was performed separately within each operating level and sensor channel to avoid confounding fault-related differences with operating-condition effects. The Kruskal-Wallis test was first used as an omnibus test to determine whether a feature differed among the represented health conditions.

The resulting p-values were corrected using the Benjamini-Hochberg false discovery rate procedure. Fault-versus-Normal pairwise comparisons were then performed using the Wilcoxon rank-sum test, again with false discovery rate correction. Pairwise comparisons were included only when both the fault and Normal groups contained at least three independent trials. Practical separation between each fault and the Normal condition was quantified using Cliff's delta:

$$\delta = \frac{1}{n_F n_N}\sum_{i=1}^{n_F}\sum_{j=1}^{n_N}\operatorname{sgn}\left(x_{F,i} - x_{N,j}\right) \quad (9)$$

where n_F and n_N are the numbers of independent fault and Normal trials, respectively. Positive values indicate higher feature values under the fault condition, whereas negative values indicate lower values. The absolute value δ was used to evaluate effect-size magnitude.

The sensor-sensitivity ranking was based on the same independent trial-level comparisons. For each fault condition, channel, feature, and represented operating level, the fault-versus-Normal effect size and FDR-adjusted p-value were calculated. For conditions represented at more than one operating level, mean absolute Cliff's delta, the number of significant operating levels, and effect-direction consistency were evaluated. Sensor–feature combinations were ranked lexicographically by all-level significance, direction consistency, number of significant operating levels, and mean absolute effect size. Conditions represented at only one operating level were labelled as not assessable for operating-level consistency. The highest-ranked sensor channel for each health condition was used only for physical interpretation and representative analytical plots. It was not used to select CNN input channels based on the unknown true label; the final CNN received all five synchronized channels.

The complete sensor-sensitivity procedure is summarized in Figure 3. Following trial-level median aggregation, each fault condition was compared with Normal operation at the same operating level and sensor channel. Statistical significance was assessed using the Wilcoxon rank-sum test with Benjamini–Hochberg FDR correction, while practical separation was quantified using Cliff's delta. Sensor–feature combinations were subsequently ranked according to statistical significance, effect-size magnitude, and operating-level consistency to identify fault-specific high-sensitivity measurement locations.

2.5 Machine-Learning Models

Six conventional supervised classifiers were initially screened using the engineered vibration features: Decision Tree, Random Forest, Support Vector Machine, k-nearest neighbour, Naïve Bayes, and a feed-forward neural network. Their principal configurations are summarized in Table 5. Preliminary screening identified the Random Forest as the conventional model selected for the final paired evaluation. The final feature-based dataset contained 948 independent trial groups and 42,660 feature observations. Each trial contributed 45 observations, corresponding to nine windows from each of the five accelerometer channels.

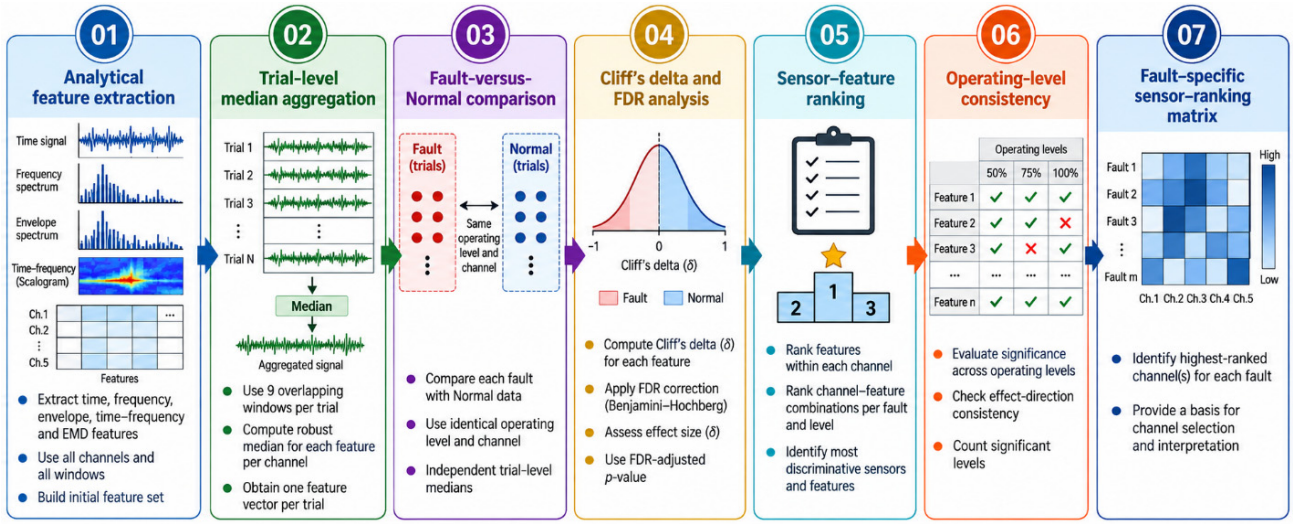


Figure 3. Workflow of the proposed trial-level sensor-sensitivity analysis and fault-specific sensor-ranking procedure.

Fixed class-stratified grouped partitioning assigned 644 trials to training, 113 trials to validation, and 191 trials to testing. This produced 28,980 training observations, 5,085 validation observations, and 8,595 test observations. All observations associated with the same TrialGroupID were kept within a single subset.

Feature standardization was performed after grouped partitioning. The mean and standard deviation of each predictor were calculated from the training observations only and then applied unchanged to the validation and testing observations:

$$z_{ij} = \frac{x_{ij} - \mu_{j,train}}{\sigma_{j,train}} \tag{10}$$

Operating level was deliberately excluded from the principal Random Forest predictor set because several health conditions were available only at particular operating levels. This reduced the possibility that the classifier would use condition–speed availability rather than vibration characteristics. The earlier speed-included model was retained only as an ablation result and was not used for the principal paired comparison.

At the observation level, each channel-window feature vector received one predicted label. At the trial level, the 45 predictions belonging to the same TrialGroupID were aggregated using majority voting. In the case of a tie, the class with the highest mean model score among the tied classes was selected. Accuracy, precision, recall, specificity, balanced accuracy, and macro F1-score were calculated using one-versus-rest class counts. Macro-averaged metrics were used to reduce the influence of class imbalance.

Table 5. Machine-learning models considered in this study

Model	Learning principle	Implemented configuration	Role in study
Decision Tree (DT)	Rule-based partitioning	Minimum leaf size = 5	Preliminary interpretable baseline
Random Forest (RF)	Bagged decision-tree ensemble	300 trees; minimum leaf size = 3; square-root predictor subsampling	Principal speed-excluded paired comparator
Support Vector Machine (SVM)	Maximum-margin classification	RBF kernel; box constraint = 1; one-vs-one ECOC	Preliminary nonlinear baseline
k-nearest neighbour (k-NN)	Distance-based classification	5 neighbours; Euclidean distance; equal weighting	Preliminary local non-parametric baseline
Naïve Bayes (NB)	Probabilistic classification	Kernel predictor distributions	Preliminary probabilistic baseline
Feedforward neural network (NN)	Multilayer nonlinear learning	Hidden layers = 32 and 16 neurons; iteration limit = 300	Preliminary feature-based neural baseline

2.6 CNN-Based Fault Diagnosis

A five-channel one-dimensional convolutional neural network was developed to classify health conditions directly from synchronized raw vibration signals. Unlike the conventional machine-learning pipeline, the CNN did not use engineered features. The complete-channel feature table was used only to identify eligible trial groups and retrieve the corresponding raw vibration recordings.

Multichannel Signal Reconstruction and Quality Control

For each eligible trial group, the original 20 kHz vibration recordings were retrieved for all five channels. A common 5 s interval was identified to ensure temporal synchronization. The channel-wise mean was removed from each signal, and the synchronized recordings were resampled to 5 kHz using the anti-aliased MATLAB resample function. Trial groups without the required five-channel common duration were excluded.

Gross-amplitude screening was then applied to remove physically unrealistic signals. A trial group was excluded if any channel contained non-finite values, an absolute peak greater than 1000 g, or an RMS value greater than 100 g. One corrupted Normal trial at the 75% operating level was removed by this procedure. The final CNN cohort contained 948 independent trials and 8,532 synchronized five-channel windows.

Each 5 s multichannel signal was segmented into 1 s windows with 50% overlap. At the resampled frequency of 5 kHz, each input window contained five channels and 5,000 temporal samples:

$$\mathbf{X}_i \in \mathbb{R}^{5\times5000} \tag{11}$$

The same grouped split used for the feature-based models was retained: 644 training trials, 113 validation trials, and 191 testing trials. These corresponded to 5,796 training windows, 1,017 validation windows, and 1,719 testing windows.

2.6.2 Normalization, Class Weighting, and Architecture

CNN normalization parameters were calculated from the training subset only. For each channel c, the mean $\mu_{c,train}$ and standard deviation $\sigma_{c,train}$ were computed using all temporal samples from the training windows. Each sample was standardized as

$$z_{i,c,n} = \frac{x_{i,c,n} - \mu_{c.train}}{\sigma_{c,train}} \tag{12}$$

The same parameters were applied unchanged to the validation and test subsets. Standardized amplitudes were clipped to the interval ([-12,12]) to reduce the influence of extreme values. Inverse-frequency class weights were calculated from the training subset only and normalized to have a mean value of one. The CNN architecture consisted of three one-dimensional convolutional blocks followed by global average pooling, dropout, a fully connected layer, softmax activation, and a weighted classification layer. The convolutional blocks used 16, 32, and 64 filters with kernel lengths of 32, 16, and 8 samples, respectively. Batch normalization and ReLU activation were applied after each convolution, and max pooling was used after the first two convolutional blocks. The complete architecture is summarized in Table 6.

Table 6. Architecture of the five-channel one-dimensional convolutional neural network

Stage	Layer	Configuration	Purpose
Input	Sequence input	Five channels × 5,000 samples; normalization disabled within the input layer	Receives a synchronized 1 s multichannel vibration window at 5 kHz
Block 1	1D convolution	16 filters; kernel length = 32; same padding	Extracts broad local temporal patterns
	Batch normalization	Applied	Stabilizes intermediate feature distributions
	ReLU	Applied	Introduces nonlinear activation
	Max pooling	Pool length = 4; stride = 4	Reduces temporal resolution
Block 2	1D convolution	32 filters; kernel length = 16; same padding	Extracts intermediate fault-sensitive patterns
	Batch normalization	Applied	Stabilizes training
	ReLU	Applied	Introduces nonlinear activation
	Max pooling	Pool length = 4; stride = 4	Further reduces temporal resolution
Block 3	1D convolution	64 filters; kernel length = 8; same padding	Learns higher-level temporal representations
	Batch normalization	Applied	Stabilizes feature distributions
	ReLU	Applied	Introduces nonlinear activation
Feature reduction	Global average pooling	One value per final feature map	Produces a fixed-length representation
Regularization	Dropout	Rate = 0.30	Reduces overfitting
Classification	Fully connected	10 output neurons	Produces one score for each health condition
	Softmax	10 class probabilities	Converts scores into class probabilities
Output	Weighted classification layer	Categorical cross-entropy with training-derived class weights	Assigns the predicted health condition

2.6.3 Network Training and Evaluation

The network was trained using the Adam optimization algorithm. The initial learning rate was 1×10^{-3}, mini-batch size of 64, maximum epoch count of 30, validation frequency of 90 iterations, and validation patience of five evaluations. The class-weighted categorical cross-entropy loss was used:

$$\mathcal{L} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{C} w_c y_{i,c} \log\left(p_{i,c}\right) \quad (13)$$

where B is the mini-batch size, C is the number of classes, w_c is the class weight, $y_{i,c}$ is the one-hot encoded target, and $p_{i,c}$ is the predicted class probability. To evaluate sensitivity to random initialization, the CNN was trained five times using random seeds 42–46. The principal CNN performance was reported as the mean and standard deviation across the five repetitions. One model was selected exclusively according to the minimum validation loss for detailed class-wise reporting and paired trial-level comparison. Test performance was not used for model selection. CNN performance was evaluated at both window and trial levels. At the window level, each synchronized five-channel segment received one predicted label. At the trial level, the nine predictions belonging to the same TrialGroupID were combined using majority voting, with summed softmax probability used to break ties. The Random Forest and CNN were finally compared on the same 191 independent test trials using exact McNemar testing. The CNN signal preparation, partitioning, and training configuration are summarized in Table 7.

Table 7. Five-channel signal preparation, grouped partitioning, and CNN training configuration

Parameter	Final setting
Original sampling frequency	20 kHz
Common synchronized duration	5 s
Channel preprocessing	Removal of channel-wise DC component
Resampling method	Anti-aliased MATLAB resample
CNN sampling frequency	5 kHz
Number of synchronized channels	5
Window duration	1 s
Samples per channel per window	5,000
CNN input dimensions	5 × 5,000
Window overlap	50%
Step length	2,500 samples or 0.5 s
Windows per independent trial	9
Initially eligible complete-channel trials	1,033
Trials excluded during raw-signal reconstruction	84
Trials excluded by gross-amplitude screening	1
Final retained trial groups	948
Total synchronized CNN windows	8,532

Parameter	Final setting
Training subset	644 trials; 5,796 windows
Validation subset	113 trials; 1,017 windows
Testing subset	191 trials; 1,719 windows
Grouping variable	TrialGroupID
Training-derived normalization	Separate mean and standard deviation for each channel
Standardized-amplitude clipping	[-12, 12]
Class weighting	Inverse-frequency weights calculated from training windows
Optimizer	Adam
Initial learning rate	1×10^{-3}
Maximum epochs	30
Mini-batch size	64
Data shuffling	Before every epoch
Validation frequency	Every 90 iterations
Validation patience	5 evaluations
Dropout rate	0.30
Training repetitions	5
Random seeds	42–46
Model-selection criterion	Minimum validation loss
Loss function	Class-weighted categorical cross-entropy
MATLAB version	R2023a
Execution hardware	CPU; Intel Core i5; 24 GB RAM

3. RESULTS AND DISCUSSION

Fault-Specific Sensor-Sensitivity Matrix

The sensor-sensitivity analysis produced the fault-specific ranking of sensor-feature combinations summarized in Table 8. The listed channels represent the highest-sensitivity measurement locations within the investigated dataset; they should not be interpreted as universally optimal sensor positions. The analytical results illustrate how the represented health conditions altered vibration amplitude, impulsiveness, spectral-energy distribution, envelope response, time-localized energy, and intrinsic-mode-function content at the selected high-sensitivity channels.

3.2 Analytical Signal-Analysis Results

Figures 5–9 present representative analytical vibration responses obtained using the fault-specific high-sensitivity sensor channels identified in Table 8. These representative signals are used to illustrate condition-dependent behaviour in the time, frequency, power-spectral-density, envelope, and time–frequency domains.

3.2.1 Time-Domain Characteristics

The representative time-domain waveforms show that the investigated health conditions modify the vibration response in different ways depending on the fault mechanism and the selected high-sensitivity sensor channel. The Normal condition exhibits a comparatively stable oscillatory response, whereas several fault conditions show increased amplitude fluctuation, intermittent peaks, or more irregular waveform behaviour. Bearing-related and looseness-related conditions are expected to produce impulsive or intermittent responses because localized contact defects and structural looseness introduce repeated impacts into the vibration signal. Impeller damage and rotor-related faults can also alter the waveform amplitude because hydraulic disturbance, imbalance, or periodic rotational excitation changes the dynamic loading of the pump–motor system. These waveform-level differences support the use of energy- and impulsiveness-related descriptors such as RMS, peak-to-peak value, kurtosis, and crest factor in the subsequent statistical and machine-learning analyses.

Table 8. Highest-ranked sensor–feature combinations based on trial-level effect-size analysis

Health condition	Ch.	Sensor location	Highest-ranked feature	Mean $\lvert\delta\rvert$	Mean signed δ	Operating-level evidence
Bearing BPFI	1	Motor NDE horizontal	Crest factor	1.000	+1.000	1/1; 100%; single level
Bearing BPFO	5	Pump NDE vertical	Spectral centroid	0.785	-0.785	2/2; 50, 100%; significant and directionally consistent
Bearing fault	4	Pump DE horizontal	Shape factor	0.841	+0.841	2/2; 50, 100%; significant and directionally consistent
Impeller damage	5	Pump NDE vertical	RMS	0.973	-0.973	2/2; 75, 100%; significant and directionally consistent
Loose foot	1	Motor NDE horizontal	RMS	1.000	-1.000	1/1; 75%; single level
New motor	1	Motor NDE horizontal	RMS	1.000	-1.000	1/1; 75%; single level

Health condition	Ch.	Sensor location	Highest-ranked feature	Mean |δ|	Mean signed δ	Operating-level evidence
baseline						
Rotor fault	1	Motor NDE horizontal	RMS	1.000	+1.000	2/2; 50, 100%; significant and directionally consistent
Soft foot	1	Motor NDE horizontal	Kurtosis	1.000	-1.000	2/2; 75, 100%; significant and directionally consistent
Stator fault	4	Pump DE horizontal	Shape factor	0.947	+0.947	2/2; 75, 100%; significant and directionally consistent

Note. NDE = non-drive end; DE = drive end; RMS = root mean square; δ = Cliff's delta. Positive signed values indicate higher feature values under the fault condition than under Normal operation; negative values indicate lower values. The operating-level evidence column reports significant/represented levels, represented levels, and consistency status.

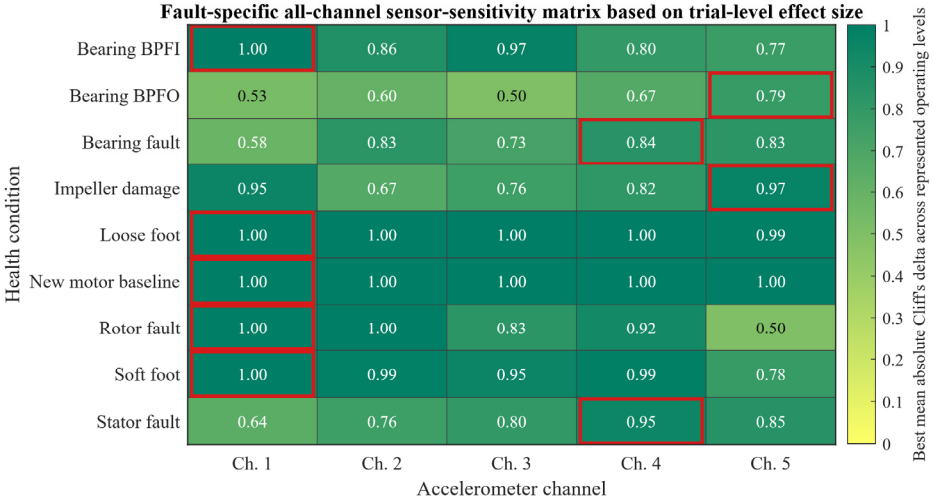


Figure 4. Fault-specific all-channel sensor-sensitivity matrix based on independent trial-level fault-versus-Normal effect sizes. Red outlines indicate the highest-ranked channel(s) for each condition.

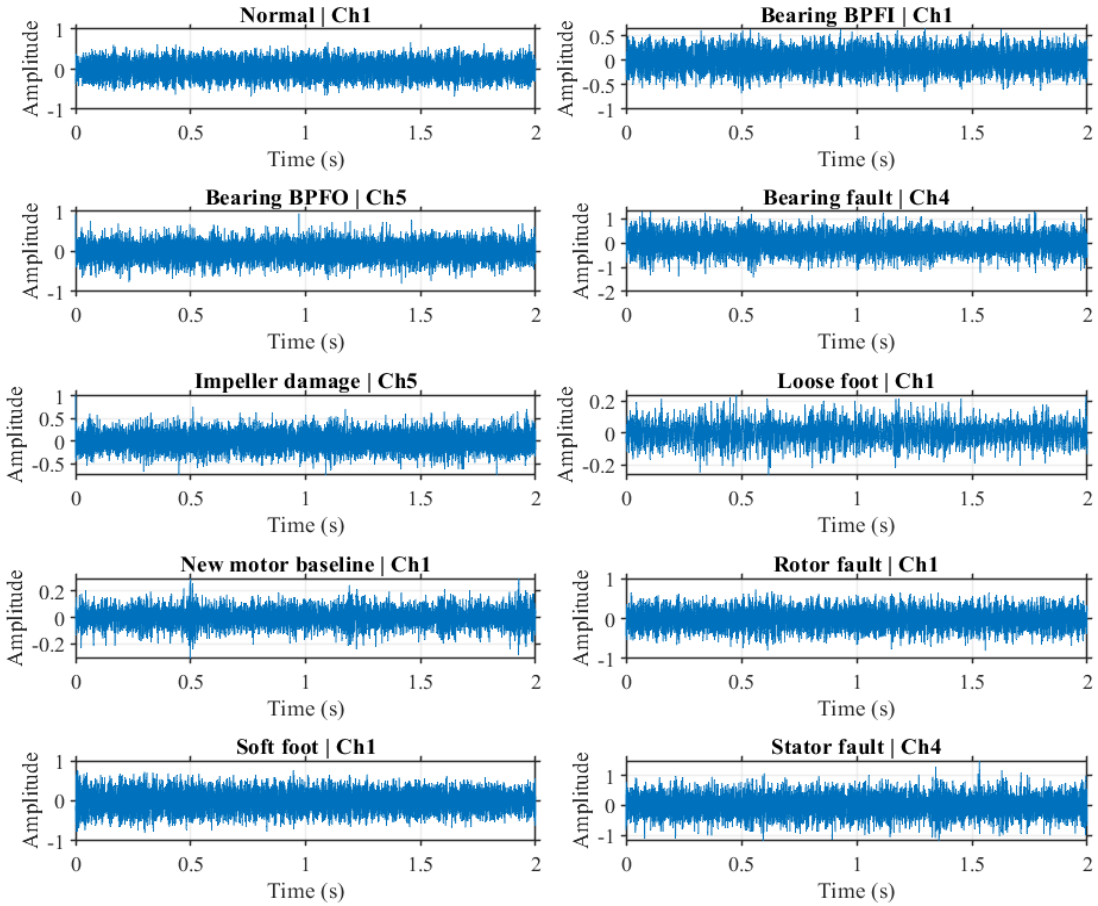


Figure 5. Representative time-domain vibration waveforms obtained from the recommended sensor channels for the investigated health conditions.

3.2.2 Frequency-Domain Characteristics

The frequency-domain characteristics of the represen–tative vibration signals were analysed using the fast Fourier transform (FFT), as shown in Figure 6. Compared with the normal operating condition, the investigated faults exhi–bited noticeable changes in the distribution of vibration energy across the frequency spec–trum. Bearing-related faults, including BPFI, BPFO, and general bearing defects, showed increased spectral activity compared with the hea–lthy condition. The presence of additional frequency com–ponents indicates the generation of repetitive mechanical impacts caused by localized defects on bearing elements.

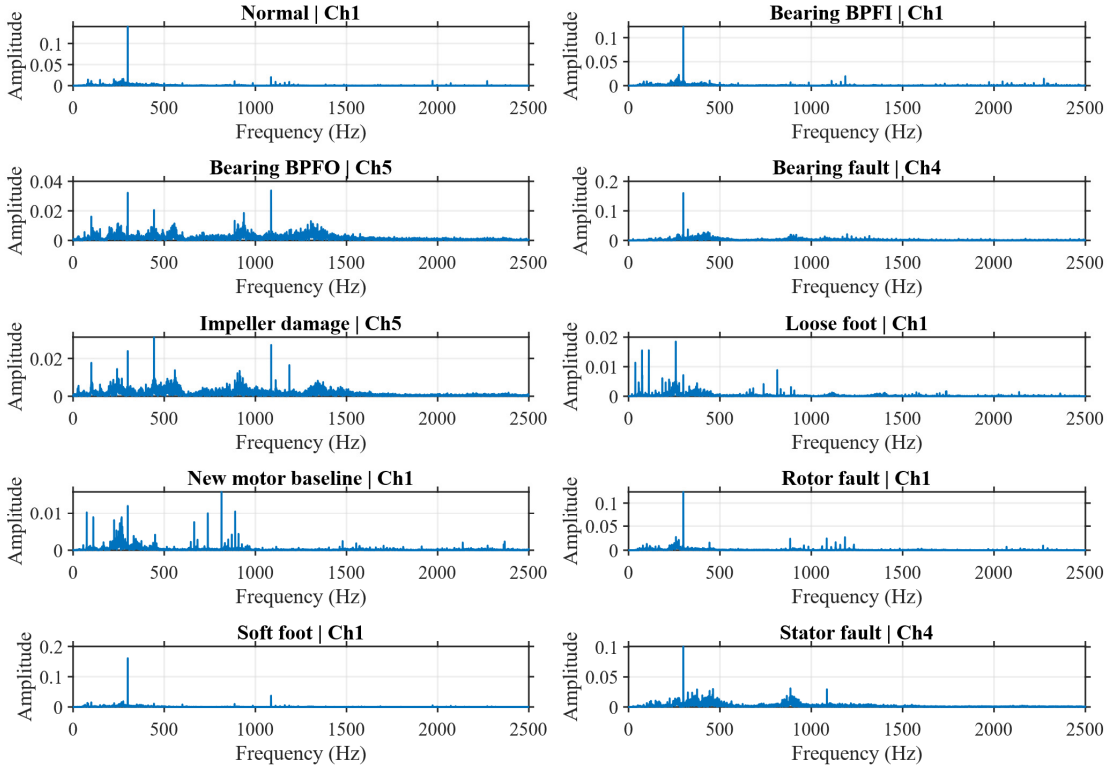


Figure 6. Frequency spectra of representative vibration signals obtained from the recommended sensor channels.

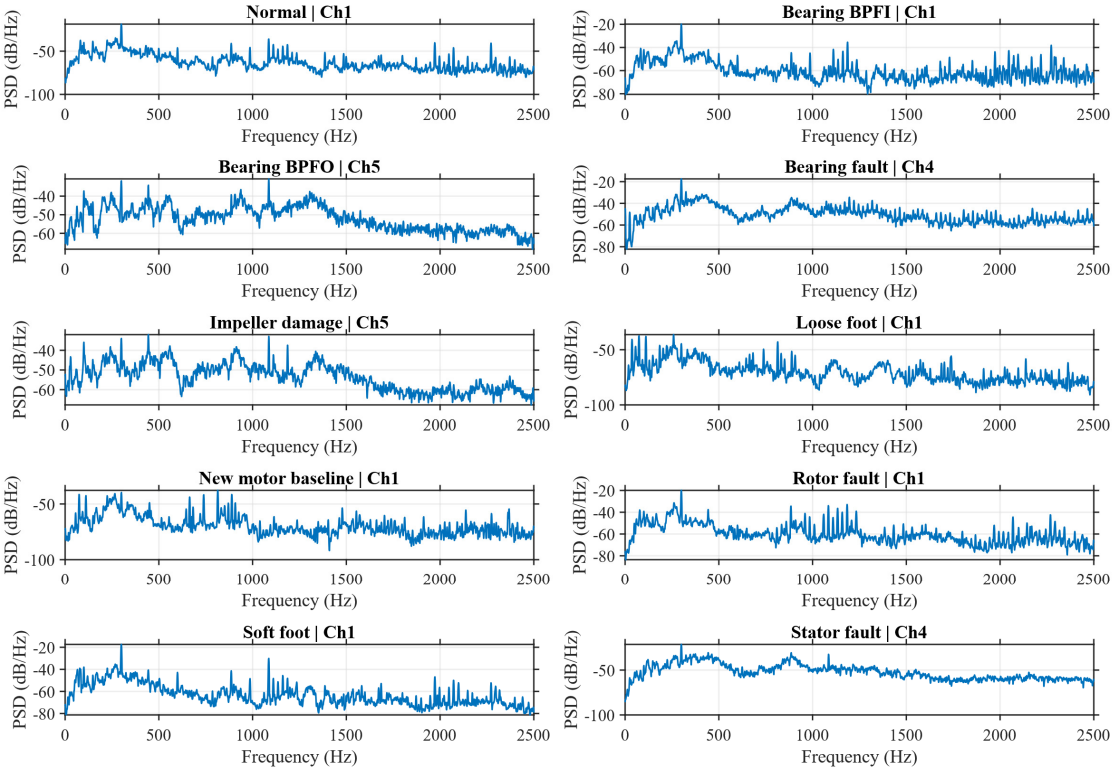


Figure 7. Welch power spectral density estimates of representative vibration signals obtained from the recommended sensor channels.

Impeller damage resulted in increased spectral amplitudes due to hydraulic imbalance and disturbance-induced vibration. Similarly, looseness-related faults produced broader frequency responses caused by nonlinear mechanical interactions and variation in structural stiffness. Rotor faults exhibited distinct frequency components associated with periodic rotational excitation, whereas stator faults demonstrated changes in spectral energy distribution due to electromechanical disturbances. These observations indicate that different failure mechanisms introduce characteristic changes in the frequency-domain response. Although frequency-domain analysis provided useful diagnostic information, overlapping spectral components among different faults highlight the requirement for additional analysis methods such as envelope analysis, time–frequency representation, and machine learning-based classification. The PSD analysis shown in Figure 7 further illustrates the redistribution of vibration energy caused by different faults. Normal operation exhibited relatively concentrated vibration energy, whereas fault conditions resulted in increased energy over wider frequency ranges. Bearing faults demonstrated enhanced high-frequency energy contribution due to impact-induced excitation. In contrast, impeller and rotor-related abnormalities showed stronger energy concentration at lower-frequency regions corresponding to mechanical excitation sources. The PSD results confirm that fault occurrence changes both the magnitude and frequency distribution of vibration energy, supporting the suitability of spectral features such as dominant frequency, spectral centroid, spectral entropy, and band-power ratios for automated diagnosis.

3.2.3 Envelope Analysis

Envelope analysis was used to highlight amplitude-modulated components that may be less obvious in the conventional frequency spectrum. Such modulation is physically meaningful in rotating machinery because localized defects, looseness, and intermittent mechanical contact can generate repeated impacts that modulate the vibration carrier signal. The envelope spectra indicate that the different health conditions produce distinct modulation-energy patterns across the analysed frequency range, although the dominant components are not identical for all faults. Bearing-related conditions generally show stronger modulation behaviour associated with repetitive defect interactions, while looseness, rotor, and impeller-related abnormalities may introduce broader or lower-frequency modulation patterns due to structural and hydraulic excitation mechanisms. These observations show that envelope-domain information provides complementary diagnostic evidence rather than replacing time- or frequency-domain analysis.

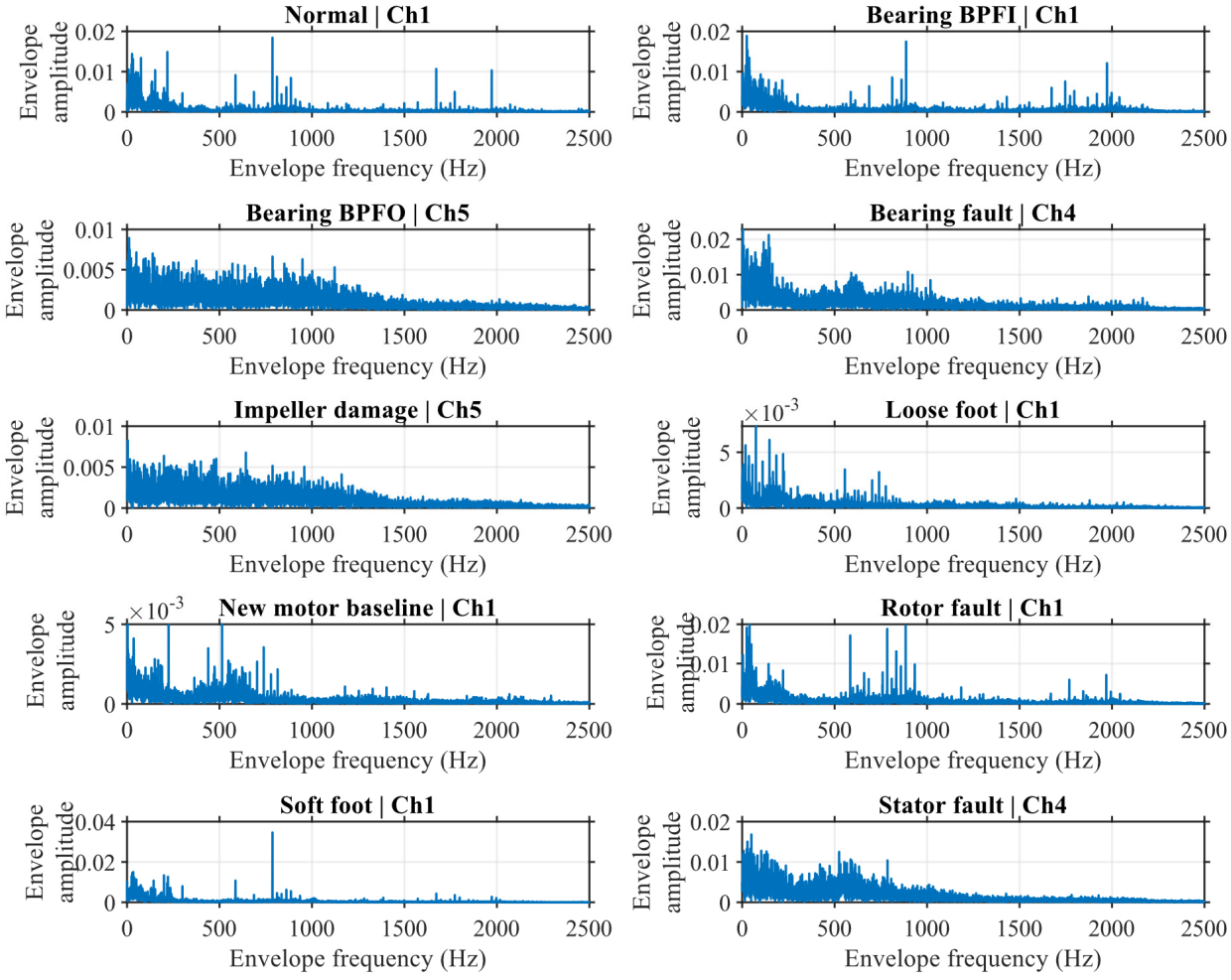


Figure 8. Envelope spectra of representative vibration signals obtained from the recommended sensor channels.

3.2.4 Time–Frequency Analysis

Figure 9 presents the STFT representations of the rep–resentative vibration signals obtained from the recom–mended sensor channels. Unlike the stationary FFT and PSD representations, the spectrograms retain the tem–poral evolution of spectral energy and therefore reveal changes in vibration content over the analysed interval. Differences in the distribution and persistence of time-localized frequency components among the investigated health conditions provide complementary diagnostic information to the time-, frequency-, and envelope-domain analyses.

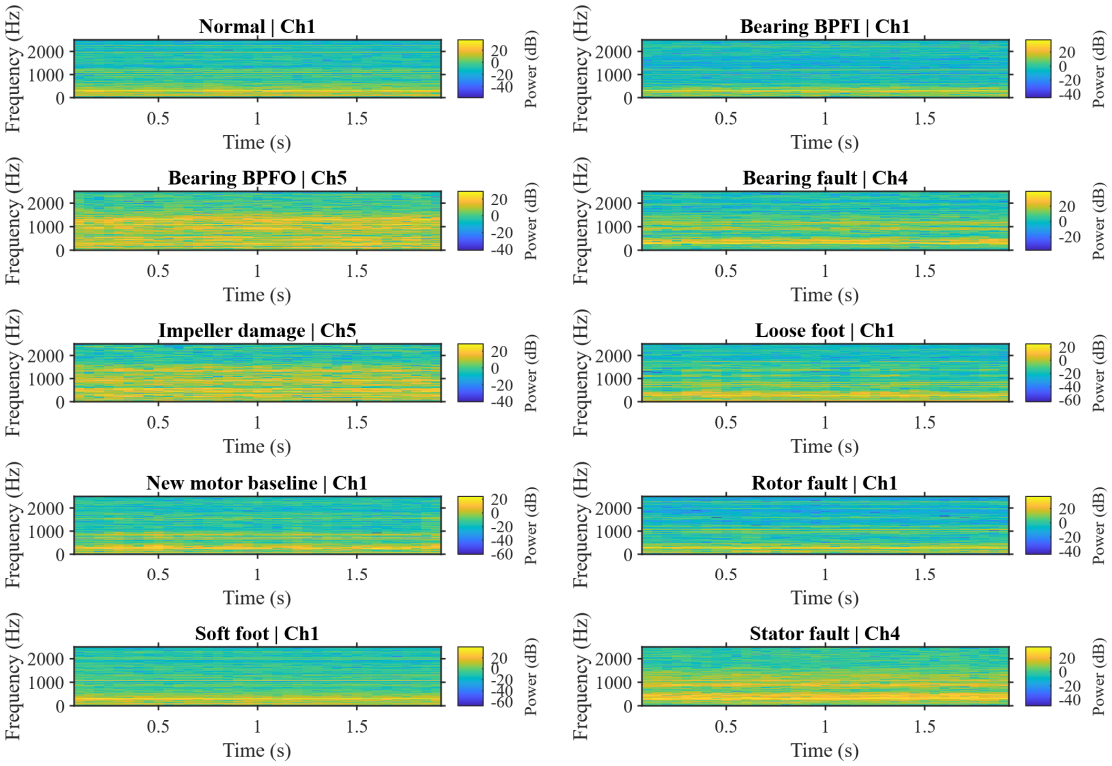
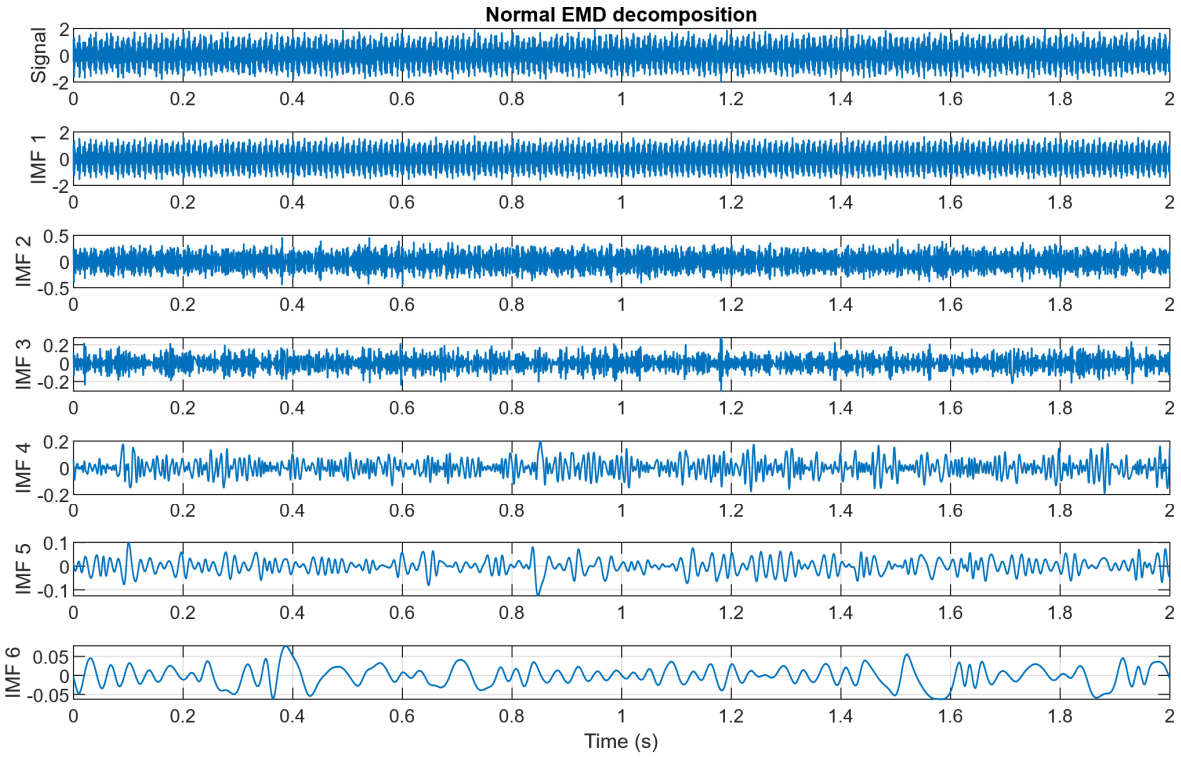
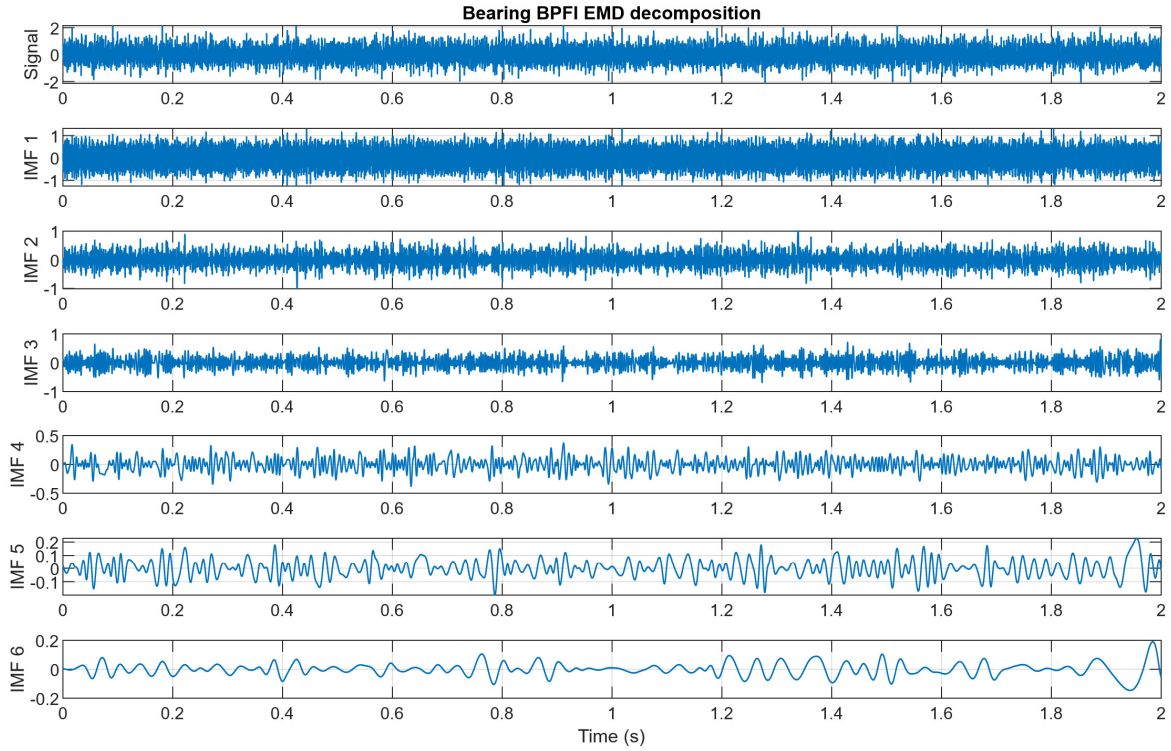


Figure 9. STFT spectrograms of representative vibration signals obtained from the recommended sensor channels.

3.2.5 Empirical Mode Decomposition



(a)



(b)

Figure 10. Representative EMD results for (a) Normal operation and (b) Bearing BPFI, showing the original signal and first six IMFs.

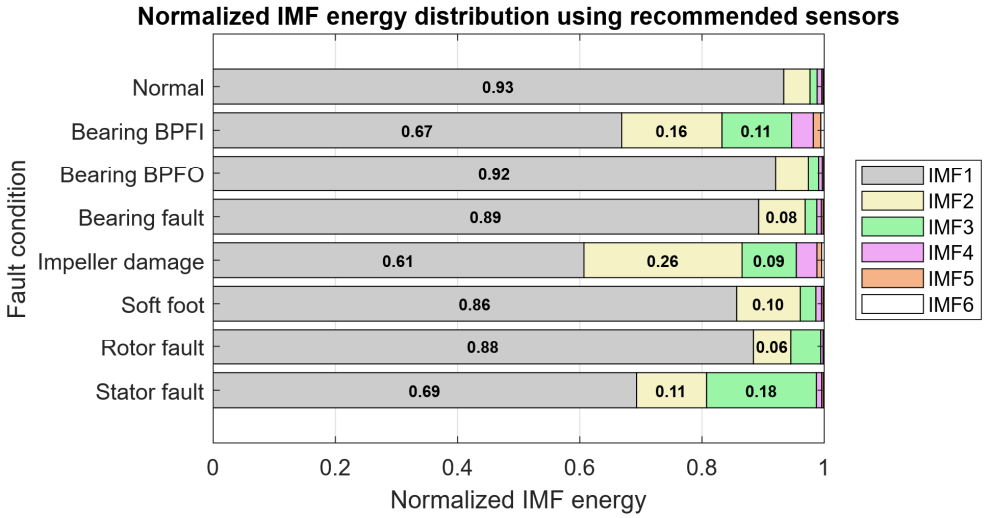


Figure 11. Normalized energy distributions of the first six IMFs for representative vibration signals.

The Normal condition exhibits relatively stationary oscillatory components with lower impulsiveness across the intrinsic mode functions (IMFs). In contrast, the Bearing BPFI signal produces pronounced high-frequency oscillatory IMFs with stronger transient components, illustrating the capability of EMD to isolate localized fault-related vibration signatures. Figure 10 compares the EMD of representative Normal and Bearing BPFI signals through the original waveform and first six IMFs. The decomposition separates oscillatory components across different time scales, enabling differences in high-frequency and transient vibration content to be examined. The corresponding normalized IMF-energy distributions are presented in Figure 11 to illustrate how vibration energy is partitioned among the decomposed modes.

3.2.6 Trial-Level Statistical Validation and Effect-Size Results

Table 9. Trial-level statistical-validation design and audit summary

Item	Final trial-level statistical design
Independent trials	948
Trial-channel summaries	4,740 (948 trials × 5 channels)
Windows aggregated per trial-channel	9; median aggregation
Feature count	21 analytical features
Omnibus analysis	Kruskal-Wallis within operating level and channel
Possible omnibus tests	315 (3 levels × 5 channels × 21 features)
Pairwise analysis	Fault versus Normal within the same level and channel

Item	Final trial-level statistical design
Multiple-testing control	Benjamini-Hochberg FDR, α = 0.05
Minimum formal pairwise sample size	At least 3 trials in each group

The trial-level analysis showed that 74.4% of eligi–ble fault-versus-Normal comparisons remained signifi–cant after Benjamini-Hochberg correction, while 67.4% exhi–bited medium or large absolute Cliff's delta effects. Spec–tral band-power fractions and impulsiveness-related desc–riptors produced the strongest overall separation. However, effect direction and magnitude varied with health condition, operating level, and sensor chan–nel, confirming that no single feature provided universal discrimination.

Table 10. Aggregate trial-level significance and effect-size outcomes

Aggregate outcome	Result	Interpretation
FDR-significant fault-versus-Normal comparisons	74.4%	Most eligible comparisons retained statistical evidence after correction
Medium or large absolute Cliff's delta effects	67.4%	A majority of comparisons showed practically meaningful separation
Most informative feature families	Spectral band-power fractions and impulsiveness-related features	Discriminative mechanisms differed among health conditions and operating levels
Operating-level consistency	Assessed only when a condition was represented at two or more levels	Single-level conditions were labelled not assessable

These findings support the use of complementary time-domain, spectral, and envelope-domain predictors in the subsequent classifiers.

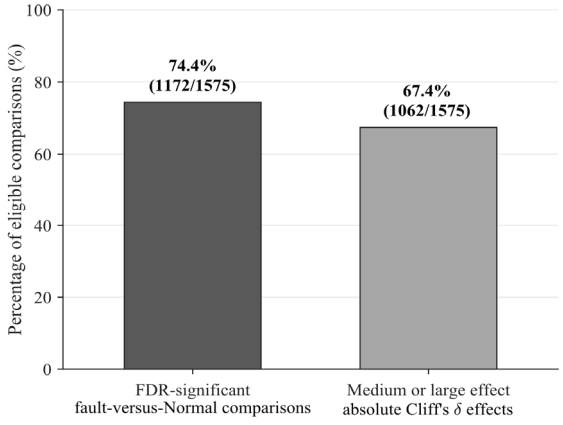


Figure 12. Aggregate outcomes of the trial-level fault-versus-Normal statistical analysis.

3.3 Machine Learning-Based Fault Classification Results

Feature-based classifiers were developed using the final common cohort of 948 independent trials and 42,660 feature observations. Each trial contributed 45 obser–vations, corresponding to nine windows from each of the five accelerometer channels. The fixed paired parti–tion comprised 644 training trials (28,980 observations), 113 validation trials (5,085 observations), and 191 tes–ting trials (8,595 observations).

Table 11. Class-wise distribution of the common trial-grouped cohort used for paired machine-learning and CNN development

Health condition	Training	Validation	Testing	Total
Bearing BPFI	31	5	9	45
Bearing BPFO	70	12	20	102
Bearing fault	102	18	30	150
Impeller damage	51	9	15	75
Loose foot	20	4	6	30
New motor baseline	10	2	3	15
Normal	266	46	80	392
Rotor fault	24	4	7	35
Soft foot	30	6	9	45
Stator fault	40	7	12	59
Total	**644**	**113**	**191**	**948**

All observations from the same TrialGroupID rema–ined within one subset. The six classifiers were used for preliminary conventional-model screening with the configurations reported in Table 5.

The Random Forest was retained as the principal conventional model because it provided the strongest validation observation-level performance while also achieving complete validation trial-level classification. It was subsequently rerun with operating level excluded from the predictor matrix. The class-wise distribution of the shared grouped cohort is presented in Table 11. The common test subset contained 191 independent trials representing all ten health conditions. At the lower level, the Random Forest classified 8,595 channel-win–dow feature observations. At the trial level, 45 obser–vation predictions were aggregated for each experiment using majority voting. This design permitted direct trial-level comparison with the five-channel CNN while preserving the different lower-level input representa–tions of the two pipelines.

3.3.1 Speed-Excluded Random Forest Performance

The speed-excluded Random Forest was trained using 644 trials, evaluated during development using 113 validation trials, and finally tested on 191 independent trials. On the common test cohort, it achieved an observation-level accuracy of 93.24%, balanced accuracy of 86.89%, macro precision of 91.90%, and macro F1-score of 88.98%. Majority-vote aggregation across the 45 channel–window observations belonging to each trial improved the trial-level accuracy to 98.95%. The corresponding trial-level balanced accuracy and macro F1-score were 95.83% and 97.11%, respectively. The trial-level result corresponds to 189 correctly classified trials out of 191. Validation performance remained strong: observation-level accuracy was 93.39%, balan–ced accuracy was 88.63%, macro F1-score was 90.06%, and all 113 validation trials were correctly classified after aggregation. Table 12 summarizes the validation and independent-test performance of the principal speed-excluded Random Forest. Results from the other preliminary classifiers were not used for the final paired inference because they were not rerun under the same speed-excluded predictor policy.

Table 12. Performance of the speed-excluded Random Forest on the validation and common independent-test cohorts

Subset	Observation accuracy	Observation balanced accuracy	Observation macro F1	Trial accuracy	Trial balanced accuracy	Trial macro F1
Validation	93.39%	88.63%	90.06%	100.00%	100.00%	100.00%
Test	93.24%	86.89%	88.98%	98.95%	95.83%	97.11%

3.3.2 Class-wise Diagnostic Performance

Class-wise trial evaluation showed that the speed-excluded Random Forest correctly classified 189 of the 191 independent test trials. The trial-level balanced accuracy of 95.83% was lower than the overall accuracy of 98.95%, indicating that the small number of errors affected class-wise recall more strongly than the aggregate accuracy. Observation-level errors also remained, as reflected by the 93.24% accuracy and 88.98% macro F1-score. The trial aggregation result should therefore be interpreted as robust experiment-level diagnosis rather than perfect classification of every constituent channel–window observation.

3.4 Deep Learning-Based Fault Classification Results

A one-dimensional convolutional neural network (1D-CNN) was developed to evaluate automatic feature learning capability directly from raw vibration signals. Unlike conventional machine learning approaches, which require manually extracted statistical and spectral features, the CNN architecture received segmented vibration windows as input and automatically learned discriminative representations from the temporal signal patterns. To ensure unbiased evaluation, the same independent trial-level grouped validation strategy was adopted. Therefore, vibration windows generated from the same experimental trial were assigned exclusively to either training, validation, or testing subsets, preventing overlap-induced information leakage.

Operating-condition-specific evaluation of the selected five-channel 1D-CNN showed complete classification of the available independent test observations at all three operating levels. At the 50% operating level, the test subset contained 531 windows from 59 independent trials representing four health-condition classes. The 75% subset contained 261 windows from 29 trials representing six classes, while the 100% subset contained 927 windows from 103 trials representing eight classes. Window- and trial-level accuracy, balanced accuracy, macro precision, macro recall, macro specificity, and macro F1-score were 100% for each operating-level subset. No prediction was assigned to a class absent from the corresponding true-label subset.

3.4.1 1D-CNN Model Performance

The five-channel CNN used the same grouped assignments shown in Table 11: 644 training trials, 113 validation trials, and 191 testing trials. These groups produced 5,796, 1,017, and 1,719 synchronized five-channel windows, respectively. The selected sensor-sensitivity rankings were used for physical interpretation only; the final CNN received all five synchronized channels and did not select an input channel using the unknown true label. Across five independent training repetitions, the five-channel 1D-CNN achieved a mean window-level accuracy of 99.76 ± 0.26%, balanced accuracy of 99.92 ± 0.06%, and macro F1-score of 99.72 ± 0.23%. Trial-level aggregation produced a mean accuracy of 99.90 ± 0.23% and a mean macro F1-score of 99.91 ± 0.21%. The repeated-run statistics are summarized in Table 14. The model selected solely by minimum validation loss classified all 1,719 test windows and all 191 independent test trials correctly. The selected-run result was retained for detailed class-wise reporting, while the corresponding confusion matrices were omitted because all predictions were correct and the matrices were fully diagonal. The small standard deviations across the five independent repetitions indicate limited run-to-run variability within the evaluated grouped test cohort.

3.4.2 Class-Wise CNN Performance

The CNN selected exclusively according to the minimum validation loss correctly classified all 1,719 synchronized five-channel test windows and all 191 independent test trials. Therefore, the selected model achieved 100% accuracy, balanced accuracy, macro precision, macro recall, macro specificity, and macro F1-score on the common test cohort. Because every prediction was correct, the corresponding confusion matrices were fully diagonal and provided no additional diagnostic information beyond the numerical results. They were therefore omitted. The selected-run result should be interpreted together with the repeated-training performance reported in Table 14 and Figure 14, which captures the small but non-zero variability across the five independent training repetitions.

3.4.3 Comparison Between Machine Learning and Deep Learning Approaches

Open circles represent the performance obtained in each independent training repetition, while filled diamonds and error bars show the mean and standard deviation across five runs.

The expanded vertical scale is used to display the small between-run variability. For the strictly paired comparison, the speed-excluded Random Forest and the five-channel CNN used identical 644/113/191 trial assignments. The Random Forest achieved 93.24% observation accuracy, 86.89% observation balanced accuracy, and 88.98% observation macro F1 on the common test cohort. After majority voting, it correctly classified 189 of the 191 independent test trials, corresponding to 98.95% trial accuracy, 95.83% balanced accuracy, and 97.11% macro F1.

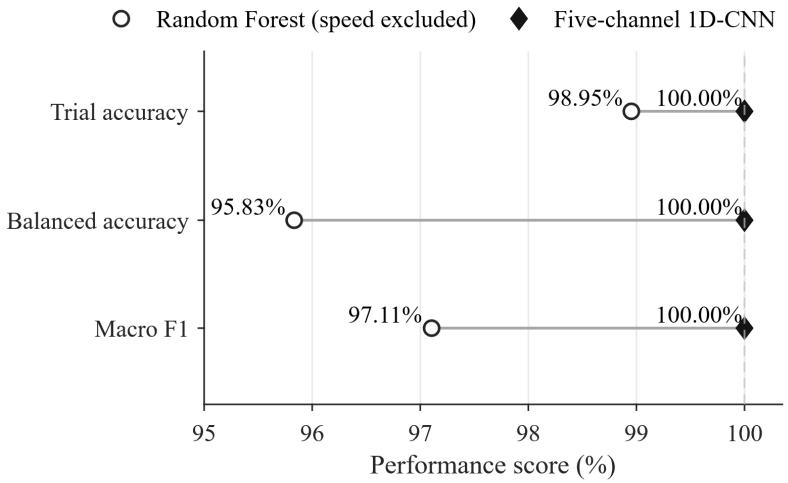


Figure 13. Paired trial-level performance of the speed-excluded Random Forest and five-channel 1D-CNN on 191 independent test trials.

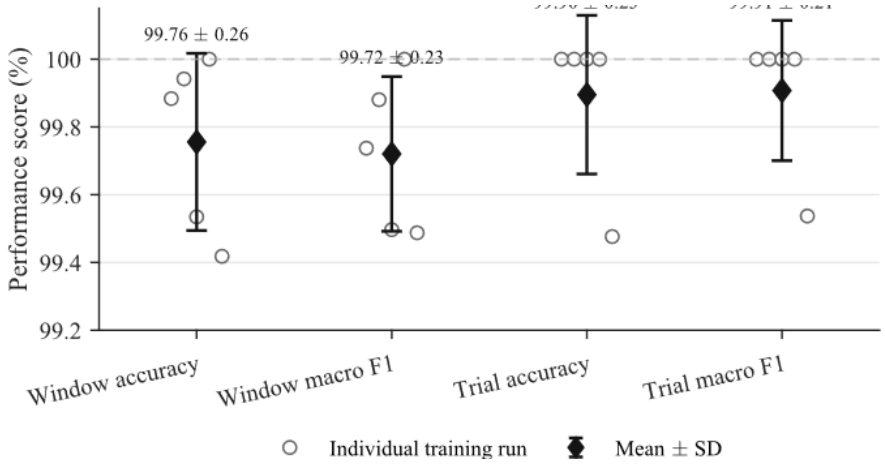


Figure 14. Performance variability of the five-channel 1D-CNN across five independent training repetitions.

The validation-selected CNN classified all 191 test trials correctly. The paired predictions therefore contained two trials that were incorrect for the Random Forest but correct for the CNN, and no trial that was correct for the Random Forest but incorrect for the CNN. The exact two-sided McNemar test produced $p = 0.500$. Thus, the CNN achieved a small numerical trial-level advantage, but the difference was not statistically significant on the available test cohort. The lower-level metrics were not directly compared because the Random Forest classified channel–window feature vectors, whereas the CNN classified synchronized five-channel windows.

The selected repeat, chosen solely by minimum validation loss, achieved 100% window- and trial-level test performance.

Table 13. Paired comparison of the validation-selected Random Forest and five-channel 1D-CNN on the same independent test trials

Approach	Input representation	Training trials	Validation trials	Testing trials	Trial accuracy	Balanced accuracy	Macro F1
Random Forest	Engineered vibration features and channel identifier; operating level excluded	644	113	191	98.95%	95.83%	97.11%
Five-channel 1D-CNN	Synchronized raw five-channel vibration windows	644	113	191	100.00%	100.00%	100.00%

Table 14. Repeated five-channel 1D-CNN performance under grouped trial-level validation

Metric	Mean ± SD
Window accuracy	99.76 ± 0.26%
Window balanced accuracy	99.92 ± 0.06%
Window macro precision	99.54 ± 0.38%
Window macro recall	99.92 ± 0.06%
Window macro specificity	99.97 ± 0.03%
Window macro F1	99.72 ± 0.23%
Trial accuracy	99.90 ± 0.23%
Trial macro F1	99.91 ± 0.21%

Table 15. Approximate computational-cost comparison between the speed-excluded Random Forest and five-channel 1D-CNN

Model	Input representation	Approximate training time	Approximate test/inference time	Hardware	Practical comment
Speed-excluded Random Forest	Engineered vibration features and channel identifier; operating level excluded	Short; typically within minutes after feature extraction	Very low; trial-level prediction from 8,595 test observations completed rapidly	MATLAB R2023a; CPU; Intel Core i5; 24 GB RAM	Computationally efficient and physically interpretable, but requires prior feature extraction
Five-channel 1D-CNN	Synchronized raw five-channel vibration windows, (5 x 5000) samples per window	Longer; repeated training required substantially higher CPU time than RF	Low after training; 1,719 test windows classified rapidly	MATLAB R2023a; CPU; Intel Core i5; 24 GB RAM	Avoids manual feature extraction and learns multichannel temporal patterns, but has higher training cost

Note. The comparison reports relative computational cost under the CPU-based MATLAB environment used in this study. Exact wall-clock time depends on MATLAB configuration, processor load, memory availability, and implementation details. Therefore, the values are reported as approximate practical timing categories rather than hardware-independent benchmarks.

3.4.4 Computational Cost and Practical Deployment Considerations

The feature-based Random Forest and the five-channel 1D-CNN differed not only in input representation and diagnostic performance but also in computational demand. The Random Forest used engineered analytical features and therefore required feature extraction before classification, but its model training and inference were comparatively lightweight. In contrast, the CNN avoided manual feature extraction and learned directly from synchronized raw vibration signals, but required longer training time because convolutional filters were optimized iteratively over multichannel time-series windows.

Both models were trained and evaluated on the same CPU-based MATLAB R2023a environment reported in Table 7, using an Intel Core i5 workstation with approximately 24 GB RAM. The comparison indicates that the Random Forest is more suitable when low computational cost and interpretability are priorities, whereas the CNN is attractive when synchronized multichannel raw signals are available and automatic representation learning is desired.

3.5 Overall Discussion and Practical Implications for Industrial Pump Condition Monitoring

The proposed framework integrates sensor-sensitivity assessment, analytical vibration interpretation, trial-level statistical validation, feature-based machine learning, and multichannel deep learning within a unified diagnostic methodology. The results demonstrate that vibration responses depend strongly on both fault mechanism and measurement location, indicating that sensor placement significantly influences diagnostic performance. The identified high-sensitivity channels should therefore be interpreted as dataset-specific recommendations rather than universally optimal sensor locations.

3.5.1 Relationship Between Analytical Features and Fault Mechanisms

The analytical investigation showed that different health conditions produced distinct responses in the time, frequency, envelope, time–frequency, and empirical-mode domains. Energy-related descriptors such as RMS and peak-to-peak amplitude reflected changes in vibration intensity, whereas kurtosis, crest factor, and envelope features highlighted impulsive behaviour. Spectral centroid, entropy, and band-power fractions characterized the redistribution of vibration energy across the frequency spectrum. Trial-level statistical analysis confirmed that many features retained significant fault-versus-Normal separation after false-discovery-rate correction; however, both statistical significance and effect size depended on operating level, sensor location, and fault type. These findings indicate that reliable diagnosis requires complementary analytical features rather than reliance on a single descriptor.

3.5.2 Comparison of Analytical, Machine Learning, and Deep Learning Approaches

Analytical signal processing provided physically interpretable evidence, whereas the Random Forest and 1D-CNN automated multiclass fault classification. The speed-excluded Random Forest correctly classified 189 of 191 independent test trials using engineered vibration features, while the selected CNN correctly classified all 191 trials using synchronized raw vibration signals. Although the CNN achieved a small numerical improvement, the paired McNemar test ($p = 0.500$) did not demonstrate statistically significant superiority. The two approaches therefore offer complementary advantages: the Random Forest provides transparent feature-based interpretation, whereas the CNN automatically learns multichannel temporal representations without manual feature engineering. The effectiveness of statistical vibration features with conventional classifiers has also been demonstrated by Samal et al. [43], who compared multiple machine-learning algorithms for cylindrical-roller-bearing fault classification. Importantly, sensor-sensitivity analysis was used only for interpretation and future instrumentation planning and was not used to select CNN inputs according to the unknown fault label.

3.5.3 Importance of Grouped Validation

Reliable performance estimation depends strongly on the validation strategy. Random window-level partitioning

may introduce similarity bias because overlapping win–dows from the same experimental trial can appear in both training and testing subsets. To avoid this problem, all windows belonging to the same TrialGroupID were assig–ned exclusively to one subset before model development. Consequently, model performance was evaluated using independent experimental trials, providing a more realistic estimate of diagnostic capability within the represented operating conditions. Nevertheless, this grouped validation does not establish generalization to unseen machines, operating levels, or industrial environments.

3.5.4 Practical Implications

The proposed framework has several practical implica–tions for intelligent pump condition monitoring. Fault-specific sensor ranking may support future sensor-placement optimization and reduced-sensor monitoring strategies, although this study did not evaluate the effect of sensor removal on diagnostic performance. The feature-based Random Forest offers physically interp–retable diagnostic decisions, while the multichannel CNN demonstrates the potential of synchronized raw vibration signals for automated fault recognition. The applicability of vibration-feature-based machine lear–ning to rotating-machinery condition assessment has also been demonstrated by Jebur and Soud [44], who evaluated several conventional classifiers for journal-bearing wear diagnosis. From an engineering perspec–tive, the con–dition-specific sensor rankings may guide the selection and placement of permanent accelero–meters during instrumentation planning, while the fea–ture-based model can support interpretable diagnostic screening when engineered vibration descriptors are available. In mainte–nance applications, the resulting diagnostic information could support fault screening and prioritization of further inspection, whereas automated multichannel monitoring may be considered when synchronized raw vibration acquisition is available. However, practical deployment will require further investigation of real-time com–putational performance, latency, sensor drift, industrial noise, and fault-progression behaviour before implemen–tation in predictive-maintenance systems.

3.5.5 Limitations and Future Work

This study was conducted using laboratory data from a single centrifugal pump configuration. Health condi–tions were not uniformly represented across operating levels, and several trial groups were excluded during multichannel reconstruction and quality-control proce–dures. The evaluation did not include unseen machines, unseen operating levels, external datasets, long-term sensor degradation, compound faults, or real-time dep–loyment. Although operating level was excluded from the principal Random Forest to reduce shortcut learning, the remaining conventional classifiers were not retra–ined using the same predictor policy. Future work sho–uld extend the proposed framework through cross-mac–hine and cross-dataset validation, leave-one-operating-level-out evaluation, uncertainty estimation, early -stage and compound-fault diagnosis, online deployment, and multimodal fusion of vibration, current, pressure, temperature, and acoustic measurements using the same leakage-controlled trial-level validation strategy.

Comparison with Previous Studies

Recent studies have demonstrated substantial impro–vements in centrifugal pump fault diagnosis using ad–vanced signal processing, feature engineering, and deep learning. However, differences in datasets, operating conditions, and validation strategies make direct accu–racy comparisons difficult. Ullah et al. demonstrated the effectiveness of wavelet-coherence representations combined with principal component analysis for vibra–tion-based pump fault classification [40]. Hu et al. improved coupled-fault diagnosis using Boruta-SHAP feature selection and an enhanced stacked denoising autoencoder, highlighting the importance of feature selection and model interpretability [41]. Koç and Saçaklıdır showed that optimized ensemble learning us–ing acoustic signals provides an alternative sensing strategy for centrifugal pump diagnosis [42].

Compared with these studies, the present work emphasizes methodological rigor rather than classifi–cation accuracy alone. The proposed framework syste–matically evaluates sensor sensitivity before analytical interpretation, employs grouped trial-level validation to reduce overlap-induced similarity bias, and integrates multi-domain vibration analysis with both feature-based and raw-signal-based intelligent models. This combi–nation provides physically interpretable diagnostic evidence together with robust automated classification, making the framework more suitable for practical labo–ratory-scale condition monitoring and providing a stronger foundation for future industrial validation.

Table 16. Comparison of the proposed framework with previous pump fault-diagnosis studies

Study	Signal/Input	Method	Main focus	Limitation addressed in present work
[40]	Vibration-based representations	Wavelet coherence + PCA	Signal representation and feature extraction	Adds sensor sensitivity evaluation and grouped validation
[41]	Extracted features	Boruta-SHAP + enhanced SDAE	Feature selection and explainability	Adds signal-level analysis and sensor-ranking guidance
[42]	Acoustic signals	Optimized ensemble learning	Alternative sensing-based diagnosis	Provides vibration-based fault-specific sensor selection
Present study	Raw vibration + analytical features	Time/Frequency/Envelope/STFT/EMD + ML + 1D-CNN	Integrated interpretation and paired trial-level evaluation	Leakage-controlled grouping, sensor-sensitivity ranking, and paired ML-CNN comparison

4. CONCLUSIONS

This study developed an integrated vibration-based framework for centrifugal-pump fault diagnosis by combining sensor-sensitivity assessment, analytical vibration interpretation, trial-level statistical validation, feature-based machine learning, and multichannel deep learning. The framework was evaluated using the publicly available University of Twente Motor-2 dataset under grouped trial-level validation to prevent information leakage between training and testing data.

The principal conclusions are summarized as follows:

1. Sensor sensitivity depended strongly on fault mechanism and measurement location. Different accelerometer channels were identified as the most informative for different health conditions, providing dataset-specific guidance for sensor placement and future reduced-sensor monitoring studies.
2. Analytical vibration analysis successfully characterized fault-dependent behaviour in the time, frequency, envelope, time–frequency, and empirical-mode domains. Trial-level statistical validation confirmed that 74.4% of eligible fault-versus-Normal comparisons remained significant after false-discovery-rate correction, while 67.4% exhibited medium or large Cliff's delta effect sizes, demonstrating the practical discriminative capability of the extracted vibration features.
3. The speed-excluded Random Forest provided strong and physically interpretable diagnostic performance, whereas the five-channel 1D-CNN achieved highly accurate classification directly from synchronized raw vibration signals. Although the CNN correctly classified all independent test trials and showed a small numerical advantage over the Random Forest, the paired McNemar test ($p = 0.500$) indicated that this improvement was not statistically significant. The two approaches therefore provide complementary benefits, combining interpretable engineered features with automatic multichannel representation learning.

Overall, the proposed framework demonstrates that integrating leakage-aware grouped validation, sensor-sensitivity assessment, statistical effect-size analysis, analytical vibration interpretation, and intelligent classification provides a rigorous and interpretable methodology for laboratory-scale centrifugal-pump fault diagnosis. The results are applicable to independent trials drawn from the represented health conditions and operating levels but should not be interpreted as evidence of generalization to unseen machines or industrial environments.

The present study is limited to one laboratory test rig, a finite range of operating conditions, and a partially unbalanced dataset. Future work should therefore investigate cross-machine and cross-domain validation, leave-one-operating-level-out evaluation, optimized sensor configurations, early-stage and compound-fault diagnosis, uncertainty estimation, multimodal sensor fusion, and real-time implementation for industrial predictive-maintenance applications.

REFERENCES

[1] C. E. Sunal, V. Dyo, and V. Velisavljevic, "Review of Machine Learning Based Fault Detection for Centrifugal Pump Induction Motors," IEEE Access, vol. 10, pp. 71344–71355, 2022, doi: 10.1109/ACCESS.2022.3187718.

[2] N. Dutta, P. Kaliannan, and S. Paramasivam, "A comprehensive review on fault detection and analysis in the pumping system," International Journal of Ambient Energy, vol. 43, no. 1, pp. 6878–6898, Dec. 2022, doi: 10.1080/01430750.2022.2056917.

[3] W. Jiang et al., "Fault diagnosis and status evaluation of axial piston pumps: Review and development prospects," Journal of Mechanical Science and Technology, vol. 40, no. 3, pp. 1795–1815, Mar. 2026, doi: 10.1007/s12206-026-0128-x.

[4] D. A. Rodrigues et al., "Fault diagnosis of electric submersible pumps using vibration signals," Journal of the Brazilian Society of Mechanical Sciences and Engineering, vol. 45, no. 9, p. 445, Sep. 2023, doi: 10.1007/s40430-023-04370-z.

[5] F. M. Varejão, L. H. Sousa Mello, M. Pellegrini Ribeiro, T. Oliveira-Santos, and A. Loureiros Rodrigues, "An open source experimental framework and public dataset for vibration-based fault diagnosis of electrical submersible pumps used on offshore oil exploration," Knowl. Based. Syst., vol. 288, p. 111452, Mar. 2024, doi: 10.1016/j.knosys.2024.111452.

[6] Z. Ahmad, A. Rai, A. S. Maliuk, and J.-M. Kim, "Discriminant Feature Extraction for Centrifugal Pump Fault Diagnosis," IEEE Access, vol. 8, pp. 165512–165528, 2020, doi: 10.1109/ACCESS.2020.3022770.

[7] S. Cao, Z. Hu, X. Luo, and H. Wang, "Research on fault diagnosis technology of centrifugal pump blade crack based on PCA and GMM," Measurement, vol. 173, p. 108558, Mar. 2021, doi: 10.1016/j.measurement.2020.108558.

[8] Z. Ahmad, A. Rai, M. J. Hasan, C. H. Kim, and J.-M. Kim, "A Novel Framework for Centrifugal Pump Fault Diagnosis by Selecting Fault Characteristic Coefficients of Walsh Transform and Cosine Linear Discriminant Analysis," IEEE Access, vol. 9, pp. 150128–150141, 2021, doi: 10.1109/ACCESS.2021.3124903.

[9] A. Husna, K. Indriawati, and B. L. Widjiantoro, "Discriminant Feature Extraction of Motor Current Signal Analysis and Vibration For Centrifugal Pump Fault Detection," in 2021 International Conference on Instrumentation, Control, and Automation (ICA), IEEE, Aug. 2021, pp. 207–212. doi: 10.1109/ICA52848.2021.9625679.

[10] S. Saeed and B. Reddy, "Diagnostics of vibrations due to looseness fault and unbalance in rotating machinery with neural network," Diagnostyka, vol. 26, no. 4, Art. no. 2025403, 2025, doi: 10.29354/diag/211587.

[11] K. Chang and S. H. Park, "Random forest-based multi-faults classification modeling and analysis for intelligent centrifugal pump system," Journal of Mechanical Science and Technology, vol. 38, no. 1,

pp. 11–20, Jan. 2024, doi: 10.1007/s12206-023 - 1202-2.

[12] A. Saxena and S. Rani, "Adopting Support Vector Machine Algorithm for Fault Diagnosis in Polymer Impeller Centrifugal Pump," in 2024 International Conference on Computing, Sciences and Communications (ICCSC), IEEE, Oct. 2024, pp. 1–5. doi: 10.1109/ICCSC62048.2024.10830305.

[13] O. R. Shah, N. V. Sridharan, S. N. Ramasamy, and S. Vaithiyanathan, "Monitoring the condition of monoblock centrifugal pumps using vibration analysis and WiSARD classifier," Journal of the Brazilian Society of Mechanical Sciences and Engineering, vol. 47, no. 1, p. 18, Jan. 2025, doi: 10. 1007/s40430-024-05340-9.

[14] M. J. Hasan, A. Rai, Z. Ahmad, and J.-M. Kim, "A Fault Diagnosis Framework for Centrifugal Pumps by Scalogram-Based Imaging and Deep Learning," IEEE Access, vol. 9, pp. 58052–58066, 2021, doi: 10.1109/ACCESS.2021.3072854.

[15] A. E. Prosvirin, Z. Ahmad, and J.-M. Kim, "Global and Local Feature Extraction Using a Convolutional Autoencoder and Neural Networks for Diagnosing Centrifugal Pump Mechanical Faults," IEEE Access, vol. 9, pp. 65838–65854, 2021, doi: 10.1109/ACCESS.2021.3076571.

[16] S. Manikandan and K. Duraivelu, "Vibration-Based Fault Diagnosis of Broken Impeller and Mechanical Seal Failure in Industrial Mono-Block Centrifugal Pumps Using Deep Convolutional Neural Network," Journal of Vibration Engineering & Technologies, vol. 11, no. 1, pp. 141–152, Jan. 2023, doi: 10.1007/s42417-022-00566-0.

[17] D. Maroufi and A. Sadighi, "Deep Denoising Convolutional Autoencoder-Based Fault Diagnosis of A Centrifugal Pump," in 2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM), IEEE, Dec. 2023, pp. 550–555. doi: 10.1109/ICRoM60803.2023.10412587.

[18] C. V. Prasshanth, S. Naveen Venkatesh, T. K. Mahanta, N. R. Sakthivel, and V. Sugumaran, "Fault diagnosis of monoblock centrifugal pumps using pre-trained deep learning models and scalogram images," Eng. Appl. Artif. Intell., vol. 136, p. 109022, Oct. 2024, doi: 10.1016/j.engappai. 2024. 109022.

[19] C. V. Prasshanth, S. N. Venkatesh, T. K. Mahanta, N. R. Sakthivel, and V. Sugumaran, "Deep learning for fault diagnosis of monoblock centrifugal pumps: a Hilbert–Huang transform approach," International Journal of System Assurance Engineering and Management, 2024, doi: 10.1007/s13198-02402447-z.

[20] M. F. Siddique, S. Ullah, and J.-M. Kim, "A Deep Learning Approach for Fault Diagnosis in Centrifugal Pumps through Wavelet Coherent Analysis and S-Transform Scalograms with CNN-KAN," Computers, Materials & Continua, vol. 84, no. 2, pp. 3577–3603, 2025, doi: 10.32604/cmc. 2025.065326.

[21] Q. Chao, H. Gao, J. Tao, Y. Wang, J. Zhou, and C. Liu, "Adaptive decision-level fusion strategy for the fault diagnosis of axial piston pumps using multiple channels of vibration signals," Sci. China Technol. Sci., vol. 65, no. 2, pp. 470–480, Feb. 2022, doi: 10. 1007/s11431-021-1904-7.

[22] T. Zhang, K. Lv, and K. Feng, "Vibration Transfer Analysis of Unbalance-Rubbing-Bearing Fault of Centrifugal Pump Considering Measuring Point of Machine Foot," in 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), IEEE, Nov. 2022, pp. 1–10. doi: 10.1109/ICSMD57530.2022. 10058320.

[23] R. Deng, D. Chen, C. Yao, M. Shao, and D. Hu, "A multi-scale sensor importance-aware attention fusion network and its applications in fault diagnosis of centrifugal pumps and axial piston pumps," Measurement, vol. 258, p. 119315, Jan. 2026, doi: 10.1016/j.measurement.2025.119315.

[24] V. Dave, H. Thakker, and V. Vakharia, "Fault Identification of Ball Bearings using Fast Walsh Hadamard Transform, LASSO Feature Selection, and Random Forest Classifier," FME Transactions, vol. 50, no. 1, pp. 202–210, 2022, doi: 10.5937/ fme2201202D.

[25] Y. Yang et al., "DSA-MGCN: a domain adaptation-based fault diagnosis method for marine centrifugal pump under unknown operating conditions," Measurement, vol. 258, p. 119022, Jan. 2026, doi: 10.1016/j.measurement.2025.119022.

[26] D. Cabrera et al., "Adversarial Fault Detector Guided by One-Class Learning for a Multistage Centrifugal Pump," IEEE/ASME Transactions on Mechatronics, vol. 28, no. 3, pp. 1395–1403, Jun. 2023, doi: 10.1109/TMECH.2022.3226347.

[27] I. M. dos S. Varejão et al., "The similarity bias problem: What it is and how it impacts vibration based intelligent fault diagnosis," Mech. Syst. Signal Process., vol. 235, p. 112822, Jul. 2025, doi: 10.1016/j.ymssp.2025.112822.

[28] X. Yin, W. He, Q. He, D. Chen, B. Zhang, and H. Zhao, "Data imbalanced fault diagnosis of aviation fuel pumps based on adaptive weighting using an enhanced broad learning system," Measurement, vol. 257, p. 118612, Jan. 2026, doi: 10.1016/j. measurement.2025.118612.

[29] W. Liu, K. Chen, Z. Wei, H. Wang, T. Zhao, and L. Song, "A Sample Enhanced Fault Diagnosis Method for Rotating Machinery Under Class-Imbalanced Scenario," in 2025 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD), IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/ ICSMD67131.2025.11365429.

[30] S. Deng et al., "Lightweight cost-sensitive multi-expert dual knowledge transfer network for imbalanced fault diagnosis," Eng. Appl. Artif. Intell., vol. 173, p. 114336, Jun. 2026, doi: 10.1016/ j.engappai.2026.114336.

[31] D. Bromberger, A. Neuenfeldt Júnior, V. Garcia, and M. Francescatto, "Interpretable machine learning in industrial water pump failure prediction: the role of Shapley additive explanations values and data sampling techniques," Eng. Appl. Artif. Intell., vol. 165, p. 113558, Feb. 2026, doi: 10.1016/j.engappai.2025.113558.

[32] Z. Xu et al., "A digital twin system for centrifugal pump fault diagnosis driven by transfer learning based on graph convolutional neural networks," Comput. Ind., vol. 163, p. 104155, Dec. 2024, doi: 10.1016/j.compind.2024.104155.

[33] L. Jiang, H. Du, Y. Bu, C. Zhao, H. Lu, and J. Yan, "Deep learning-based multilabel compound-fault diagnosis in centrifugal pumps," Ocean Engineering, vol. 314, p. 119697, Dec. 2024, doi: 10.1016/j.oceaneng.2024.119697.

[34] D. Kumar, N. Singh Ranawat, P. Kumar Kankar, and A. Miglani, "InceptionV3 based blockage fault diagnosis of centrifugal pump," Advanced Engineering Informatics, vol. 65, p. 103181, May 2025, doi: 10.1016/j.aei.2025.103181.

[35] A. L. Dias, M. R. Buzoli, V. R. da Silva, J. C. R. da Silva, A. C. Turcato, and G. S. Sestito, "Edge-based intelligent fault diagnosis for centrifugal pumps in microbreweries," Flow Measurement and Instrumentation, vol. 101, p. 102730, Jan. 2025, doi: 10.1016/j.flowmeasinst.2024.102730.

[36] A. Saxena and S. Rani, "Vibrothermal signal-based fault detection in 3D-printed polymer pump impellers," Flow Measurement and Instrumentation, vol. 108, p. 103201, Mar. 2026, doi: 10.1016/j.flowmeasinst.2026.103201.

[37] Z. Li, X. Zhang, H. Lin, W. Qin, J. Qi, Z. Chen, and Q. Liu, "Automated Machine Learning for Fault Diagnosis Using Multimodal Mel-Spectrogram and Vibration Data," Computer Modeling in Engineering & Sciences, vol. 146, no. 2, Art. no. 16, 2026, doi: 10.32604/cmes.2026.075436.

[38] H. O. Ataç, M. Güçyetmez, and H. Bakır, "Quantifying the environmental footprint of mechanical faults: A case study on centrifugal pumps using vibration data and fault detection via machine learning methods," Progress in Engineering Science, vol. 3, no. 2, Art. no. 100298, 2026, doi: 10.1016/j.pes.2026.100298.

[39] S. Bruinsma, R. D. Geertsma, R. Loendersloot, and T. Tinga, "Motor current and vibration monitoring dataset for various faults in an E-motor-driven centrifugal pump," Data Brief, vol. 52, p. 109987, Feb. 2024, doi: 10.1016/j.dib.2023.109987.

[40] N. Ullah, M. F. Siddique, and J.-M. Kim, "A Hybrid Classification Framework of Centrifugal Pumps Using Wavelet Coherence Visuals and Principal Component Analysis," in 2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/ SmartCity /DependSys), IEEE, Dec. 2023, pp. 906–911. doi: 10.1109/HPCC-DSS-SmartCity-DependSys60770.2023.00131.

[41] K. Hu, H. Sun, W. Fan, Q. Si, Y. Wu, and S. Yuan, "Coupled fault diagnosis for centrifugal pumps through Boruta-Shap feature selection and rime-enhanced stacked denoised autoencoder," Eng. Appl. Artif. Intell., vol. 162, p. 112659, Dec. 2025, doi: 10.1016/j.engappai.2025.112659.

[42] S. Koç and I. Saçaklıdır, "Acoustic-based diagnosis of impeller and bearing faults in centrifugal pumps using optimized ensemble learning models," Journal of Mechanical Science and Technology, vol. 40, no. 1, pp. 117–131, Jan. 2026, doi: 10.1007/s 12206-025-1209-y.

[43] P. K. Samal, R. Srinidhi, I. M. Jamadar, and V. Baligar, "A Comparative Analysis of Machine Learning Algorithms for Fault Classification in Cylindrical Roller Bearings," FME Transactions, vol. 53, no. 4, pp. 681–692, 2025, doi: 10.5937/fme2504681K.

[44] N. A. Jebur and W. A. Soud, "Wear Fault Diagnosis in Journal Bearings Using Vibration Analysis and AI," FME Transactions, vol. 53, no. 3, pp. 401–408, 2025, doi: 10.5937/fme2503401J.

NOMENCLATURE

1D-CNN	One-dimensional convolutional neural network
AC	Accelerometer channel
Adam	Adaptive moment estimation
BPFI	Ball pass frequency inner race
BPFO	Ball pass frequency outer race
CNN	Convolutional neural network
CF	Crest factor
DE	Drive end
DT	Decision Tree
ECOC	Error-correcting output codes
EMD	Empirical mode decomposition
FFT	Fast Fourier transform
FDR	False discovery rate
GMM	Gaussian mixture model
IMF	Intrinsic mode function
k-NN	k-nearest neighbour
ML	Machine learning
NB	Naïve Bayes
NDE	Non-drive end
NN	Neural network
PCA	Principal component analysis
PSD	Power spectral density
RF	Random Forest
RBF	Radial basis function
ReLU	Rectified linear unit
RMS	Root mean square
SD	Standard deviation
SHAP	Shapley additive explanations
STFT	Short-time Fourier transform
SVM	Support vector machine

ДИЈАГНОСТИКА КВАРОВА ЦЕНТРИ–ФУГАЛНИХ ПУМПИ У ВИШЕ РАДНИХ УСЛОВА УЗ УЗИМАЊЕ У ОБЗИР ЦУРЕЊА, ПРИМЕНОМ КАРАКТЕРИСТИКА ВИБРАЦИЈА ОСЕТЉИВИХ НА СЕНЗОРЕ, АЛГОРИТМА „RANDOM FOREST" И ВИШЕКАНАЛНЕ 1D-CNN МРЕЖЕ

С. Саед, Б. Реди

Поуздана дијагностика кварова центрифугалних пумпи захтева прецизну класификацију поткреп–љену доказима који су физички тумачиви. Ова сту–дија представља интегрисани оквир заснован на анализи вибрација, који обједињује аналитичку обраду сигнала, анализу осетљивости сензора, стат–истичку валидацију на нивоу појединачних мерења (проба), машинско учење засновано на карактерис–тикама и петоканалну једнодимензионалну конво–луциону неуронску мрежу (1D-CNN). Скуп података „Motor-2" Универзитета у Твентеу, који обухвата десет стања исправности и три нивоа рада, анали–зиран је коришћењем пет канала акцелерометара. Након претходне обраде и контроле квалитета, 948 независних мерења искоришћено је за развој модела. Статистичка анализа је показала да је 74,4% поре–ђења стања квара у односу на нормално стање остало статистички значајно након Бењамини-Хох–берг корекције, док је 67,4% показало ефекте средње или велике величине према Клифовој делта (Cliff's delta) метрици. Током пет понављања обуке, 1D-CNN мрежа је постигла просечну тачност од 99,76 ± 0,26% на нивоу временског прозора и 99,90 ± 0,23% на нивоу појединачног мерења. На 191 независном тест-мерењу, модел „Random Forest" (без укључи–вања података о бр–зини) постигао је тачност од 98,95%, док је одабрани CNN модел исправно кла–сификовао сва мерења. Предложени оквир интег–рише тумачиву анализу вибрација са робусним вишеканалним дубоким учењем ради поуздане дијагностике кварова центрифугалних пумпи.