electronics

MDPI

Article

Signal Processing Optimization in Resource-Limited IoT for Fault Prediction in Rotating Machinery

Robertas Ūselis †, Artūras Serackis * and Raimondas Pomarnacki *,†

Department of Electronic Systems, Faculty of Electronics, Vilnius Gediminas Technical University, Plytinės g. 25, P2 234., LT-10223 Vilnius, Lithuania; robertas.uselis@vilniustech.lt

* Correspondence: arturas.serackis@vilniustech.lt (A.S.); raimondas.pomarnacki@vilniustech.lt (R.P.)

† These authors contributed equally to this work.

Abstract

Traditional fault detection methods, often designed for centralized or cloud-based systems, are ill-suited for the edge. The deployment of predictive maintenance solutions on ultra-low-cost embedded platforms remains a significant challenge due to strict limitations in memory, processing capacity, and energy availability. To address these challenges, vibration and motor current signals were analyzed using an ultra-low-cost RP2040 microcontroller. For fault detection, this study uses statistical time-domain features and principal component analysis (PCA), followed by classification with eXtreme Gradient Boosting (XGBoost) models distilled for resource-constrained deployment. Experimental evaluation demonstrated that vibration-based features achieved a diagnostic accuracy of 94.1%, while current-based representations obtained 95.5% accuracy when using principal components, compared to 83.2% with handcrafted statistical features. Model distillation reduced memory footprint by up to 2.5× (from 0.42 MB to 0.15 MB) without compromising diagnostic fidelity, enabling deployment within the 264 KB RAM and 2 MB Flash constraints of the RP2040 microcontroller. This study proposes a modular framework that systematically evaluates statistical features, dimensionality reduction, sensor synchronization, and model distillation, thereby identifying the most cost-efficient combination of techniques that balances diagnostic accuracy with strict memory and processing constraints. The findings establish that accurate fault detection can be realized directly on severely resource-limited hardware, thereby extending the practical applicability of condition monitoring to cost-sensitive industrial environments.

Keywords: Internet of Things (IoT); Predictive Maintenance (PdM); Edge Computing; Motor Current Signature Analysis (MCSA); Vibration Monitoring; Principal Component Analysis (PCA); XGBoost

check for updates

Academic Editor: Ping-Feng Pai

Received: 21 August 2025
Revised: 11 September 2025
Accepted: 12 September 2025
Published: 17 September 2025

Citation: Ūselis, R.; Serackis, A.; Pomarnacki, R. Signal Processing Optimization in Resource-Limited IoT for Fault Prediction in Rotating Machinery. *Electronics* **2025**, *14*, 3670. https://doi.org/10.3390/electronics14183670

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The increasing availability of low-cost Internet of Things (IoT) hardware has enabled the deployment of predictive maintenance systems directly at the edge of industrial processes. This trend has opened up new possibilities for real-time monitoring [1] and analysis of machinery health, potentially reducing downtime and maintenance costs. However, the use of resource-constrained devices presents unique challenges for implementing effective fault detection algorithms. Condition monitoring of electric motors relies on the accurate interpretation of time-series signals, typically acquired from current sensors and accelerometers. These signals can be analyzed in the time domain, frequency domain, or

time–frequency domain [2], each offering diagnostic perspectives and computational trade-offs [3]. Feature extraction techniques can be broadly categorized into linear and non-linear methods. Linear techniques, such as principal component analysis (PCA), aim to reduce dimensionality by projecting correlated variables into orthogonal components, thereby filtering noise while preserving the most informative structure of the data [4]. Non-linear methods, on the other hand, often rely on kernel transformations or manifold learning to capture more complex relationships at the cost of higher computational overhead.

Advanced microprocessors and external server infrastructures are commonly employed in modern fault diagnosis systems to meet the high computational demands of data-driven approaches. These include machine learning and deep learning algorithms applied to high-bandwidth sensor data, such as vibration, current, thermal, or acoustic signals, for real-time prognostics and predictive maintenance [5]. Despite notable progress in machine learning-based fault detection, most existing approaches assume access to high-performance hardware or cloud infrastructure, which is impractical in cost-sensitive industrial settings. Ultra-low-cost microcontrollers, are typically regarded as insufficient for reliable diagnostic pipelines. This study challenges that assumption by investigating whether accurate and noise-robust fault detection can be executed directly on such resource-constrained platforms.

Key diagnostic modalities for electrical motor fault detection include electromagnetic field monitoring, temperature sensing, infrared thermography, radio-frequency emission analysis, acoustic and vibration monitoring, chemical analysis, and Motor Current Signature Analysis (MCSA), alongside AI-based inference (e.g., neural networks and fuzzy logic) [6]. Typical faults exhibit distinct signatures: bearing faults induce amplitude and phase modulations in current and vibration that can be summarized by time-domain statistics—RMS, Kurtosis, and Crest Factor—as proxies of severity [7]; broken-rotor-bar faults introduce asymmetries that generate sidebands and envelope amplitude modulations in the stator current, often observable as shifts in RMS and reductions in crest factor [1]; voltage unbalance drives current asymmetry and thermal stress that scale with the voltage–imbalance ratio and frequently co-manifest with elevated mechanical vibration [8]; and mechanical misalignment is a primary source of excessive vibration, typically exciting the third harmonic (3×) in both vibration and current spectra [9]. In practice, multiple defects may co-occur, and their non-linear coupling complicates measurement and diagnosis; recent biologically inspired approaches address this challenge by coupling spiking neural networks with wavelet-based pooling (wavelet gradient) to extract discriminative time–frequency features for compound-defect detection [10]. Compound faults, resulting from the combination of multiple single faults, introduce non-linear interactions that obscure characteristic fault signatures and lead to diagnostic ambiguity, making them significantly more challenging to identify than single faults. Reinforcement learning combined with adversarial domain adaptation enables the recognition of compound faults even when only single-fault labels are available [11]. To address the reliability assessment challenge with multi-component failure dependence under small sample conditions, Pend et al. proposed a method based on uncertainty distribution and the R-vine copula model. The method accurately represents the statistical characteristics of component lifetimes, flexibly constructs the dependencies between different component failures [12].

The scope of this work extends beyond the evaluation of a single classification algorithm, addressing instead the broader challenge of enabling reliable fault detection on resource-constrained platforms. The study examines how diagnostic fidelity can be maintained under strict hardware limitations by simultaneously reducing the feature space to its most discriminative components, tailoring classification models to the constraints of embedded deployment, and ensuring coherent integration of heterogeneous sensor signals.

Within this framework, statistical features and PCA-derived representations of vibration and current signals are coupled with distilled XGBoost classifiers optimized for the RP2040 microcontroller (Raspberry Pi Ltd., Cambridge, UK). The resulting pipeline employs compact feature sets, dimensionality reduction, and rigorous memory management, thereby demonstrating that real-time and accurate (>95%) fault classification of rotating machinery is feasible even under severe computational and cost constraints of ultra-low-cost hardware. Furthermore, the study introduces a modular platform that integrates dimension reduction in input signals with the distillation of machine learning classifiers, providing a systematic basis for scalable and resource-efficient fault detection solutions.

Several microcontroller platforms can be used for edge-based signal processing, including RP2040, STM32, ESP32, and ATSAMD21 [13,14]. These devices vary in terms of power consumption, floating-point support, peripheral availability, and cost. Recent studies on embedded predictive maintenance frequently rely on the ESP32 microcontroller (Espressif Systems, Shanghai, China) due to its integrated floating-point unit, wireless connectivity, and relatively generous memory resources. However, reliance on ESP32 introduces a baseline of hardware capabilities, including floating-point arithmetic and up to 520 KB of SRAM, which cannot be assumed in the lowest-cost industrial scenarios. Moreover, the ESP32 is less suitable as a modular solution, since its integrated peripheral set limit flexibility when tailoring the system to different deployment requirements. In contrast, the RP2040 features a dual-core architecture with deterministic timing, flexible programmable I/O (PIO), and a highly predictable memory model. Despite lacking hardware floating-point support, it provides sufficient computational headroom for time-domain feature extraction when optimized integer routines are used. The RP2040 microcontroller offers a dual-core processor and programmable I/O (PIO) blocks that enable precise real-time signal acquisition and synchronization. PIO-based data handling makes it inherently better suited for multimodal fault analysis, whereas the ESP32 can suffer from jitter effects when handling multiple sensor streams. Furthermore, low price, widespread availability, and strong community support make it a compelling choice for cost-sensitive diagnostics.

2. Related Works

Basic motor fault diagnosis system can be based on an STM32 microcontroller and artificial intelligence. The system uses an ADXL335 vibration sensor to collect motor vibration data, and a Kalman filter is applied to reduce noise and improve data accuracy [15]. However, the STM32 microcontroller has limited computational resources, which makes it unsuitable for running AI models directly on the device. To optimize the data transmission to the computer for processing, salient features are first extracted from the sensor data on the microcontroller. This is achieved by transforming the filtered time-domain signals into the frequency domain using the FFT algorithm implemented on the STM32. A deep convolutional neural network (CNN) trained to classify faults is run on a more powerful host computer. When the system relies on a computer for model training and inference, it limits deployment in standalone, real-time industrial applications. To address these computational constraints, more advanced approaches utilize significantly more powerful microcontrollers. For example, Quan et al. used a much more powerful and expensive STM32H753 microcontroller [16], employing advanced time-frequency domain feature extraction methods. Their system implements Discrete Wavelet Transform (DWT) directly on the microcontroller for sophisticated multi-resolution analysis of vibration signals, capturing both transient and stationary characteristics.

However, it is noteworthy that the ESP32 microcontroller is far more prevalent in the literature for edge-AI applications. Harrabi et al. used a convolutional autoencoder model deployed on an ESP32 microcontroller for anomaly detection by reconstructing normal

daily temperature sequences. Evaluation using real-time data yielded 92% accuracy in fault detection, demonstrating the system's effectiveness. By applying quantization and pruning techniques, the model size was reduced, enabling real-time deployment on an ESP32 microcontroller with limited memory and energy capacity [17]. Their optimized model consumed only 50 mW of power and utilized 1.2 MB of memory.

To overcome the limitations of single-modal diagnostics, recent research is increasingly exploring multimodal approaches that integrate complementary sensor data [18–20]. Gana et al. further demonstrate the practical implementation of a multimodal monitoring approach by connecting both a custom lead-free piezoelectric vibration sensor and an infrared thermopile (TPA81) temperature sensor to an ESP32 microcontroller [21]. The sensors communicated with the microcontroller, showcasing the ESP32's capability to handle multiple data streams from different sensing modalities simultaneously. A lightweight artificial neural network (ANN) was then deployed directly on the ESP32 to fuse these features and diagnose imbalance faults, achieving a high accuracy of 97.69%. Microcontroller modularity allows the integration of various sensors and functionalities such as vibration, temperature, and humidity sensors can be incorporated depending on the monitoring requirements without redesigning the core architecture. This makes the proposed approach highly adaptable to underdeveloped areas [22].

However, achieving efficient and reliable fault detection requires the precise identification of the most informative features from the sensor data, which demands knowledge of which features have the greatest impact on the diagnosis. To reduce computational demands, explainable AI (XAI) techniques can be employed, aimed at making neural network decisions interpretable for humans. This approach was used to identify which input frequency components were most relevant for classification. By applying XAI-based feature selection, the input dimensionality was reduced without any performance loss, and the final classifier was retrained using only the selected frequency bands [23]. Adaptive frameworks transfer knowledge from large, cloud-trained models to compact, edge-deployable models, using techniques such as temperature scaling, multi-teacher distillation, and adversarial learning. This results in significant model compression (up to 80%), memory reduction (up to 96.58%), and inference speedup (up to 8.79×), with minimal accuracy loss [24].

Chaoraingern and Numsomran used TinyML on a Cortex-M0+ microcontroller (RP2040) to perform real-time remaining useful life (RUL) estimation of UAV Li-Po batteries. Their approach combined sensor fusion with a compact feedforward neural network optimized through Edge Impulse EON™ compiler and 8-bit quantization, enabling inference within 2 ms using less than 2 KB RAM and 11 KB Flash [25]. Compared to baseline machine learning models such as k-nearest neighbors (KNN) and random forest regression, the TinyML-based FFNN consistently achieved superior predictive accuracy. This demonstrates that resource-constrained MCU platforms can reliably support predictive maintenance tasks without relying on cloud computing. By leveraging TinyML with binary neural networks, the proposed approach reduced activation memory by nearly 28×, weight storage by 7.5×, and inference latency by 3.6×, proving that deep learning can be deployed on ultra-constrained MCUs [26].

In resource-constrained platforms, particularly those deployed for edge-based predictive maintenance, the selection of signal features must carefully balance diagnostic fidelity with computational feasibility. This trade-off is especially critical in low-power real-time systems, where algorithmic simplicity and energy efficiency directly influence the viability of the deployment [27]. Consequently, a key design principle is the identification of a minimal yet discriminative set of features that maintains high classification accuracy under strict resource constraints. The purpose of feature extraction is to reduce data dimensionality while preserving fault-related information, thereby making fault de-

tection more feasible in resource-constrained environments. As demonstrated in this study, lightweight implementations can achieve nearly indistinguishable diagnostic performance compared to their floating-point counterparts while dramatically improving computational efficiency. A good practice for practitioners operating in such environments is to start with a compact feature set and small artificial intelligence (AI) models [28], and only scale up the model complexity if the classification accuracy proves insufficient. This conservative approach reduces unnecessary resource expenditure and aligns with the iterative nature of real-world embedded diagnostics. Moreover, the ML model distillation paradigm provides a promising strategy to compress complex neural networks into significantly smaller models while preserving their predictive capacity, making them suitable for deployment in low-energy, low-memory systems [29,30]. When combined with efficient feature extraction routines, model distillation enables edge devices to benefit from high-level inference capabilities without incurring the computational cost of full-scale architectures. In this context, time-domain methods and PCA offer a pragmatic balance between diagnostic power and computational tractability, especially when compared to more memory intensive techniques. In contrast, frequency-domain techniques, such as the Fast Fourier Transform (FFT) and its variants, require the availability of a complete signal window in RAM before spectral components can be computed, thereby introducing latency and increasing memory pressure. These limitations hinder their suitability for hard real-time diagnostics, although certain optimized methods, such as the Goertzel algorithm [31], enable real-time tracking of individual frequency bins with reduced memory usage, their applicability remains narrow and often task specific. Time–frequency methods, including wavelet transforms [32], offer a compromise by capturing both temporal and spectral characteristics, but are computationally intensive.

3. Materials and Methods

In this study, the experimental setup (Table 1) was designed to analyze a three-phase induction motor operating under steady-state conditions at the nominal grid frequency of 50 Hz. The motor specifications are as follows:

Table 1. Specifications of the induction motor used in the experiments.

Parameter	Value
Motor type	Three-phase induction motor
Power rating	2.2 kW
Rated speed	1410 rpm (async) | 1500 rpm (sync)
Rated current	5.09 A
Power factor	0.81

All experimental setup, illustrated in Figure 1, consists of a low-cost and resource-constrained embedded system designed to perform on-device condition monitoring and fault classification for rotating electric machinery. The core of the system is the RP2040 microcontroller, which serves as the central processing unit, responsible for sensor data acquisition, real-time feature extraction, and inference execution.

Two sensing modalities are integrated:

- The SCT013 current sensor (YHDC, Beijing, China) is connected to the motor's power supply line and provides an analog signal proportional to the instantaneous load current. This signal is sampled by the RP2040's onboard ADC at a fixed rate (2 kHz), enabling MCSA.
- The MPU-6050 vibration sensor (InvenSense Inc., Sunnyvale, CA, USA) is mounted directly on the motor chassis to capture triaxial accelerometer data representing me-

chanical oscillations. Communication with the MPU-6050 is established via the I^2C interface, ensuring low-latency digital transfer of vibration signals.

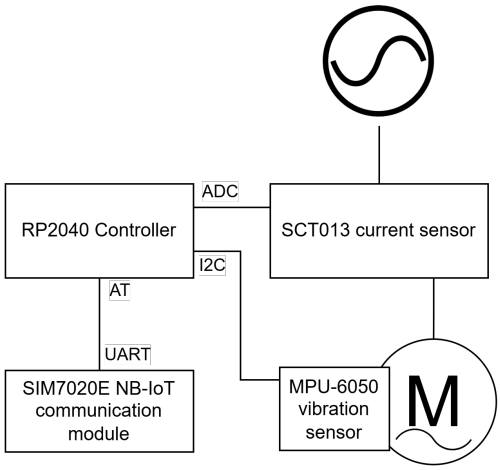


Figure 1. Experimental setup.

This study has a narrow scope to focus exclusively on the signal processing layer within the broader context of predictive maintenance systems. While end-to-end Predictive Maintenance (PdM) pipelines typically encompass sensor placement, wireless transmission, cloud integration, and decision support interfaces, these aspects lie outside the boundaries of this work. Here, we deliberately isolate the feature extraction stage, aiming to identify a minimal and computationally efficient features that retain high diagnostic discriminative power under strict memory and processing constraints. The objective is not to propose a complete system architecture, but rather to systematically research the trade-off between feature complexity and classification accuracy within embedded microcontroller environments. In this context, we treat the selection and real-time computation of statistical features as the core optimization target, assuming a fixed hardware configuration and known sensing modalities. This reductionist approach enables controlled experimentation and quantitative assessment of how specific signal descriptors contribute to accurate fault classification, forming a methodological foundation for future deployment-focused studies.

3.1. Dataset

The dataset used in this study was acquired under controlled laboratory conditions, using a well-instrumented induction motor test bench with induced fault scenarios. Each type of fault was introduced, isolated, and maintained under repeatable operating conditions, including fixed rotational speed, stable load, and minimal environmental interference. The fault conditions investigated in this work are summarized in Table 2. This setup ensures a high signal-to-noise ratio, clear fault manifestations, and minimal external variability, allowing a precise evaluation of feature effectiveness and algorithmic performance. The laboratory experiments serve as a necessary first step in isolating signal characteristics and optimizing embedded computation, and future work must extend this methodology to field deployments, incorporating non-ideal signal conditions, sensor drift, and long-term monitoring effects. While controlled laboratory datasets lack some of the variability and complexity of field data, their use is deliberate and well-suited to the objectives of this study. The primary contribution of this work is not to benchmark fault detection under uncontrolled industrial conditions, but to demonstrate that reliable classification of major motor faults is feasible on ultra-low-cost, resource-constrained hardware.

Table 2. Motor fault conditions tested.

Type	Fault	Fault Number
Mechanical	Rotor Misalignment	(4)
	Faulty Bearings	(2)
Electrical	Voltage Unbalance	(3)
	Broken Rotor Bars	(1)
Healthy	Healthy Motor	(0)

The dataset is fully balanced, each of the five fault conditions is represented by 20,000 instances, resulting in a total of 100,000 samples. Each condition corresponds to 10 s of signal acquisition at a sampling frequency of 2 kHz. Signals were segmented into non-overlapping sliding windows of 200 samples, resulting in 100 observations per second. Feature extraction was performed independently for each window, yielding a total of 100,000 instances suitable for supervised learning.

3.2. Fault Detection Methodology

To develop a diagnostic methodology that is both computationally efficient and effective in practice, it is essential to ground the design of signal processing algorithms in the capabilities and limitations of the target platform. In resource-constrained environments, this means carefully evaluating the trade-offs between various analytical approaches—not only in terms of diagnostic resolution, but also with respect to memory footprint, execution latency, and power consumption. The edge device typically computes nine time-domain features [33], including Mean, Root Mean Square, Standard Deviation, Skewness, Kurtosis, Crest Factor [34], Latitude Factor, Shape Factor (SF), and Impulse Factor [35]. These features have been shown to reduce computational complexity while maintaining effectiveness in detecting motor faults [36–38]. However, not all features contribute equally to fault diagnosis, and various methods can be employed to analyze feature significance [39–41].

Given these computational trade-offs, our experimental workflow, depicted in Figure 2, was explicitly designed around efficient time-domain analysis. The data acquisition phase involves simultaneous sampling of current and vibration signals at a rate of 2 kHz for a duration of 10 s, resulting in approximately 20,000 samples per sensor channel per fault condition. Raw signals collected from current sensors and accelerometers are initially verified for quality and then systematically stored, ensuring accurate alignment and labeling corresponding to predefined fault conditions.

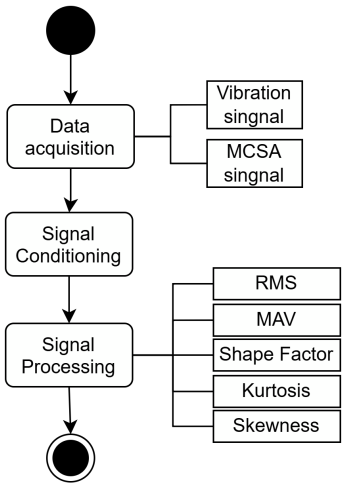


Figure 2. Workflow of the experiment.

Before feature extraction, the acquired raw signals were subjected to basic signal conditioning steps. For both current and vibration signals, a simple digital low-pass filter was applied to attenuate high-frequency noise and ensure stable statistical features in the time domain. This lightweight filtering, based on first-order IIR structure [42], was selected to balance computational simplicity with minimal signal distortion. However, detailed analysis or optimization of filter characteristics was not the focus of this study and is left for future research.

To further reduce the computation overhead, this work selects only the most significant time-domain features as identified in related studies [39,43,44], namely Mean Absolute Value (MAV), Root Mean Square, Skewness, Shape Factor (SF), and Kurtosis. This streamlined set of features ensures an optimal balance between algorithmic efficiency and diagnostic accuracy. For the most relevant time-domain features of motor vibration analysis, several studies have highlighted key statistical features and their effectiveness in fault detection and diagnosis. The most relevant time-domain features for motor vibration analysis include RMS, Peak, Kurtosis, and Skewness [45]. Crest Factor and Mean Value are also useful, though the Mean Value is less effective under varying load conditions [46]. Waveform Length, Slope Sign Changes, Zero Crossing, and Simple Sign Integral are additional features that have shown high performance in specific studies, particularly for mechanical fault identification.

For systematic identification of the minimal set of statistical features in the time domain that preserve diagnostic discriminative power under embedded constraints, we employed the minimum Redundancy Maximum Relevance (mRMR) algorithm [47]. This method balances mutual information between each candidate feature and the target class label y, while simultaneously penalizing redundancy between features. Formally, the mRMR score of a candidate subset S is defined as (1).

$$Score = \frac{1}{|S|}\sum_{f\in S} MI(f,y) \; - \; \frac{2}{|S|(|S|-1)}\sum_{i<j} MI(f_i,f_j) \tag{1}$$

where $MI(f,y)$ denotes the mutual information between feature f and the class label, while $MI(f_i,f_j)$ quantifies redundancy between feature pairs.

The feature pool comprised 15 commonly used statistical [34,45,46,48,49] features extracted in the time domain: *Mean, RMS, Root, MAV, P2P, SD, Skewness, Variance, Form factor, Kurtosis, Crest Factor, Shape Factor, MAD, Entropy, Activity.*

The second strategy for feature reduction and diagnostic insight is based on dimensionality reduction through principal component analysis (PCA). PCA is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components [4]. PCA enables the reduction in the initial feature space into a few principal components that explain the majority of data variance. Based on component loadings, the most influential factors associated with industrial motor failures are identified. This provides an empirical foundation for the early recognition of risk factors and the prioritization of maintenance decisions [50]. By selecting only the principal components with the largest eigenvalues, PCA tends to filter out noise and preserve the most significant structure in the data [4].

In this work, PCA is chosen for feature reduction due to its versatility, optimality in capturing variance, robustness in handling high-dimensional data, and the availability of efficient computational techniques. While there are other methods available, PCA's broad applicability and ease of implementation make it a preferred choice in many scenarios [51,52]. PCA represents a favorable trade-off between computational simplicity and dimensionality reduction efficiency, as its projection step reduces to a matrix–vector multiplication. In contrast, more advanced methods such as ICA, manifold learning ap-

proaches (e.g., t-SNE, UMAP), or neural autoencoders typically require substantially larger memory footprints for storing model parameters and additional runtime operations (e.g., non-linear activations, neighbor graphs). As a result, under the same SRAM budget, the maximum feasible window size on ultra-low-cost microcontrollers would be reduced even further compared to PCA. Given the strict RAM limitations, particular care was taken to ensure that the stored PCA projection matrix does not exceed 30 KB in memory, thus enabling lightweight dimensionality reduction directly on the edge hardware without runtime recomputation. This design choice reflects the practical constraints of the RP2040 microcontroller, which allocates its 264 KB SRAM across multiple tasks. It should be noted that PCA becomes impractical for large input windows on RP2040, as the memory requirements for buffering high-dimensional vectors quickly exceed the available on-chip SRAM. To determine the maximum feasible window length under a on-chip SRAM budget, we model the memory footprint of PCA inference as shown in (2).

$$M(d,k) = d\,b_{\text{in}} + d\,k\,b_W + d\,b_\mu + k\,b_W + \mathcal{O}, \tag{2}$$

where d is the window length, k the number of retained components, b_{in} the input sample width, b_W the storage width of the projection matrix, b_μ the mean vector width, and $\mathcal{O}$ the implementation overhead.

Solving $M(d,k) \leq B$ yields a closed-form upper bound for the maximum feasible input window on the RP2040. Under the configuration used in this work ($k = 4$ retained components, 16-bit input samples, quantized 16-bit projection matrix and mean vector, and a 2 KB overhead reserve), the practical limit is found to be 2388 samples per window, which at a sampling frequency of 4 kHz.

To empirically validate the diagnostic significance of the selected time-domain features and PCA, a supervised classification model was trained using the eXtreme Gradient Boosting (XGBoost) algorithm. XGBoost has emerged as a state-of-the-art method for fault detection in rotating machinery, owing to its robust handling of non-linear relationships, noise resilience, and high classification accuracy across various industrial datasets [53]. Its application in multiple studies further highlights its potential for enhancing operational reliability and reducing unexpected downtime in condition monitoring systems [54,55], with recent comparative analyses clearly demonstrating that XGBoost consistently achieves the best overall performance in terms of accuracy, precision, recall, and F1 metrics [55]. A stratified split was applied to divide the dataset into training and testing sets in a 75:25 ratio, with a fixed seed (42) to preserve reproducibility. To enable efficient deployment on resource-constrained microcontrollers, the trained XGBoost model was subsequently transformed using a tree-flattening procedure. In this process, the ensemble of decision trees is compiled into a sequence of deterministic statements, eliminating recursion and dynamic memory access. This transformation reduces execution latency, improves timing predictability, and ensures that the model can be executed on ultra-low-power embedded platforms while preserving its diagnostic validity. For the simple case with an indoor and a single operating point setup as in our dataset, classical ML methods should suffice [56], and there is no reason to use any deep leaning algorithms, especially in recourse-constrained environments.

Particular attention was given to ensuring that the resulting model remains compact enough to be deployed on a low-memory RP2040 microcontroller. Preliminary empirical constraints guided the model size target to approximately 200 KB, based on the assumption that this capacity enables sufficiently accurate fault detection while fitting within the device's Flash and RAM memory limits. However, this assumption serves as an initial baseline and does not represent a rigorously validated bound. A more comprehensive investigation is required to systematically characterize the trade-off between model

complexity, memory footprint, and classification accuracy, particularly in the context of embedded inference for predictive maintenance. In practice, the most critical constraint is the 264 KB of available SRAM, which must be shared between model execution, real-time signal buffering, and auxiliary routines such as communication and preprocessing. Therefore, the deployed model must remain sufficiently compact to preserve operational headroom, ensuring deterministic performance and preventing memory contention during runtime execution.

To fine-tune the classifier and balance its predictive performance with model size constraints, hyperparameter optimization was performed using the Optuna framework [57]. This approach allowed an efficient exploration of the parameter space through Bayesian optimization. The key parameters (Table 3) influencing both the expressiveness of the model and its memory footprint were systematically varied. Specifically, the maximum tree depth was restricted to a range of 2 to 4 levels to limit the number of nodes per tree, thereby reducing the overall model size and inference latency. The number of boosting rounds was limited to 25 to 70, ensuring that the ensemble remains sufficiently compact while still allowing gradient-based refinement. The learning rate was explored in the range between 0.01 and 0.3, balancing convergence speed and generalization capacity. This optimization strategy enabled the discovery of parameter configurations that provided high diagnostic accuracy with minimal resource overhead. Notably, the hyperparameter search space was deliberately bounded to reflect deployment constraints on the RP2040, prioritizing lightweight models suitable for inference under strict Flash and RAM limitations.

Table 3. Optuna-based hyperparameter search space for XGBoost, constrained for deployment on the RP2040 MCU.

Parameter	**Range/Values**	**Rationale for Chosen Limits (RP2040 Constraints)**
objective	multi:softprob	
eval_metric	mlogloss	
max_depth	[2, 4]	Shallow trees minimize node count, reducing RAM usage and inference latency while retaining essential non-linear capacity.
n_estimators	[25, 70]	Restricts ensemble size; more trees would improve accuracy but exceed memory and Flash limits.
learning_rate	[0.01, 0.30]	
subsample	[0.40, 1.00]	
colsample_bytree	[0.50, 1.00]	
reg_alpha (L1)	[0.00, 0.50]	
reg_lambda (L2)	[0.50, 2.00]	
Search budget	n_trials = 800	

Recent studies have shown that deep learning methods, including convolutional neural networks (CNNs) [25,58,59] recurrent neural networks such as LSTM and GRU [28], and transformer-based architectures [58], achieve state-of-the-art performance in machinery fault diagnosis. While these models provide automated feature extraction and high diagnostic accuracy, they typically require substantial computational resources (GPU/TPU) and memory footprints in the order of tens to hundreds of MB, which makes them impractical for ultra-low-cost microcontrollers such as RP2040. However, emerging TinyML techniques demonstrate that resource-optimized neural architectures can be deployed even on severely constrained MCUs through aggressive quantization (e.g., 8-bit, 1-bit),

pruning, and specialized toolchains such as TensorFlow Lite Micro or Edge Impulse EON compiler [25].

To complement deep learning approaches, we experimentally evaluated TinyML baselines in our study—specifically, a post-training quantized MLP (int8) and a binary neural network (BNN). These models remain significantly smaller in footprint (30–70 KB) compared to standard DL architectures, while still achieving competitive accuracy. The BNN consists of two hidden layers with 416 binary units (weights and activations) followed by a float32 five-way softmax output layer. The multilayer perceptrons (MLP) consist of two hidden layers with 176 ReLU-activated units and a five-neuron softmax output layer. Quantized MLP sizes correspond to post-training int8 weight quantization, whereas BNN sizes are theoretical bit-counts of the binary weights. Training was performed with a batch size of 256, a learning rate of 1×10^{-3}, and 150 epochs for BNN or 100 epochs for MLP.

Table 4 summarizes the performance of BNN and MLP trained either on raw 200-dimensional vibration windows or on 5-dimensional PCA features. Results show that raw input models achieve higher accuracy and F1 scores, but at the cost of larger model sizes, while PCA-based models are more compact yet less accurate. Notably, the int8-quantized MLP maintains nearly the same accuracy as its float counterpart, confirming the feasibility of TinyML deployment under strict memory constraints. However their diagnostic performance is sensitive to input dimensionality reduction and generally did not surpass the accuracy of our distilled XGBoost classifiers. This motivates the exploration of lightweight alternatives, such as statistical features combined with dimensionality reduction and compact ensemble models, as investigated in this work.

Table 4. Comparison of BNN and MLP (float, int8 PTQ) models on raw 200-dimensional signals vs. PCA-reduced 5-dimensional features.

Model	Accuracy	Macro-F1	LogLoss	Size (KB)
		RAW 200D		
BNN (binary)	0.744	0.749	0.647	39.43
MLP (float32)	0.803	0.784	0.352	—
MLP (int8 PTQ)	0.795	0.772	0.352	71.44
		PCA 5D		
BNN (binary)	0.317	0.296	1.330	29.52
MLP (float32)	0.707	0.694	0.678	—
MLP (int8 PTQ)	0.688	0.680	0.683	37.88

3.3. Sensor Synchronization

To ensure that the extracted features accurately represent the underlying fault dynamics, it is not sufficient to analyze each sensor modality in isolation. Reliable multimodal diagnostics additionally require precise temporal synchronization between the current and vibration signal streams. Even small cumulative desynchronization can lead to phase mismatches between modalities, degrading the correlation structure of multimodal inputs and reducing classifier accuracy. In this context, maintaining intersensor alignment below 10% of the sampling interval achieved in this setup ensures that the features derived from the current and vibration signals describe the same physical event.

To empirically ensure accurate alignment, current and vibration signals were acquired using the same RP2040 microcontroller in a synchronized fashion. The synchronization mechanism employed in this study relies on a simple software-based timer combined with shared sequential read access to both the ADC (for current) and I²C (for vibration) channels. The timing offset analysis confirmed high temporal consistency between both data streams, with a mean delay of 18.91 µs and low variance ($\sigma^2 = 4.35$). The maximum observed

offset did not exceed 35 µs, representing less than 7% of the 500 µs sampling interval at 2 kHz, well within acceptable limits for real-time multimodal signal fusion in predictive maintenance applications.

While this synchronization approach offers ease of implementation and demonstrated sufficient temporal precision for statistical feature extraction, it remains a best-effort solution that is sensitive to software jitter and I²C bus latency. More deterministic timing could be achieved by leveraging hardware timer interrupts (Timer IRQ) to trigger sensor reads, thus reducing sampling jitter and improving inter-sensor phase alignment. However, implementing such hardware-level synchronization would require careful coordination of interrupt priorities, bus contention resolution, and possibly DMA-based data handling. These aspects were deliberately excluded from the present work to maintain focus on signal processing feasibility. However, future research should investigate interrupt-driven acquisition architectures to assess their impact on diagnostic accuracy, particularly in scenarios involving higher frequency dynamics or phase-sensitive features.

4. Results

4.1. Vibration Feature Validation

To assess the diagnostic value of vibration-based features under the same fault classification framework, an additional dataset was collected using the MPU6050 inertial measurement unit. Only the embedded three-axis accelerometer was activated during data acquisition, while the gyroscope module was deliberately disabled to preserve it for future studies involving rotational dynamics or angular instability analysis. The accelerometer provided raw vibration signals along the X, Y, and Z axes, capturing both radial and axial components of structural oscillation. To reduce redundancy and improve diagnostic interpretability, a feature correlation analysis was performed using the time-domain statistics computed from the three-axis MPU-6050 accelerometer signals.

To assess statistical dependencies among the extracted vibration features, a correlation matrix was computed (Figure 3), capturing the pairwise relationships across MAV, Skewness, and Kurtosis. This analysis served to evaluate feature redundancy and inform the selection of a minimal yet informative set of descriptors suitable for embedded classification.

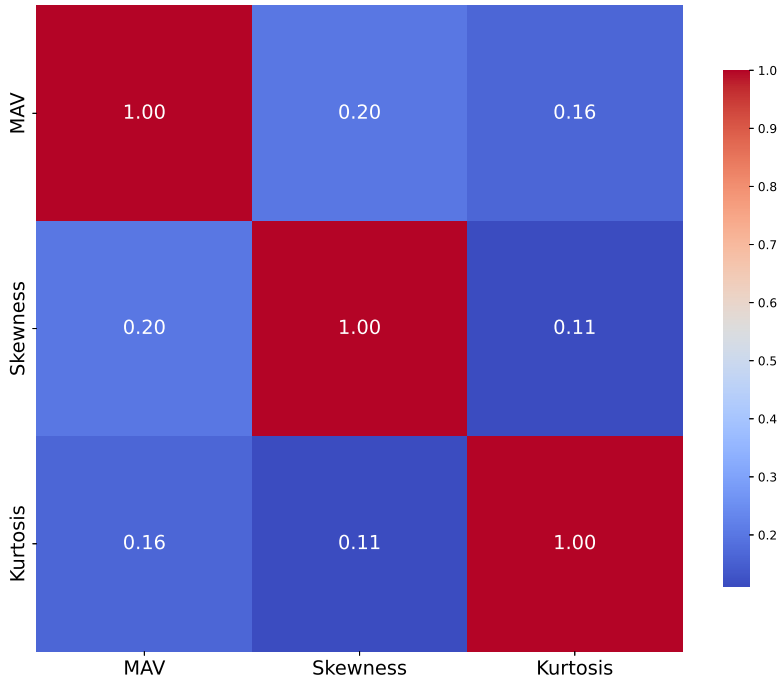


Figure 3. Vibration feature correlation matrix.

Each feature appears to capture independent aspects of the signal, suggesting strong orthogonality in their diagnostic content. Specifically, the low inter-feature dependency indicates that variations in one feature are not systematically mirrored in the others, implying that no single feature dominates or encapsulates the same statistical behavior as the rest. This distribution of information content improves the model's capacity to learn distinct fault-related patterns from the data.

The classification results presented in Table 5 demonstrate that the XGBoost model trained on unimodal statistical vibration features achieved a high level of diagnostic accuracy across all fault categories. With an overall accuracy of 94% and balanced performance metrics (precision, recall, and F1 score) for each class, the classifier exhibited consistent generalization ability without overfitting to any specific condition.

Table 5. XGBoost classification report using vibration features.

Class	Precision	Recall	F1 Score	Support
No fault (0)	0.99	0.97	0.98	75
Broken rotor bars (1)	0.95	1.00	0.97	75
Bearing fault (2)	0.91	0.83	0.87	75
Unbalance (3)	0.91	0.95	0.93	75
Misalignment (4)	0.95	0.96	0.95	75
Overall accuracy		0.94 (375 test samples)		
Macro average	0.94	0.94	0.94	375
Weighted average	0.94	0.94	0.94	375

Model size: 0.42 MB; distilled model size: 0.19 MB.

Notably, the model was able to distinguish subtle fault conditions such as unbalance and misalignment—faults that often produce low-amplitude quasi-periodic oscillations—indicating the sufficiency of time-domain statistical descriptors in capturing relevant fault characteristics. The relatively strong performance on broken rotor bar and bearing faults further reinforces the suitability of MAV, Skewness, and Kurtosis as low-complexity indicators for mechanical degradation. Importantly, the final trained model exhibited a compact memory footprint of 0.42 MB, which was further reduced to 0.19 MB after model distillation. This size is well within the flash memory constraints of edge devices such as the RP2040. For clarity, the classification metrics reported in Table 5 are defined as follows.

Accuracy measures the proportion of correct predictions out of the total predictions and is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision measures the proportion of correctly predicted positive samples out of all predicted positives and is calculated as follows:

$$Precision = \frac{TP}{TP + FP}$$

Recall (also known as sensitivity) measures the proportion of correctly predicted positive samples out of all actual positives and is calculated as follows:

$$Recall = \frac{TP}{TP + FN}$$

F1 score is the harmonic mean of precision and recall, balancing their trade-off, and is calculated as follows:

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Support refers to the number of actual occurrences of each class in the dataset and is expressed as an integer count:

$$Support = \text{Number of samples per class}$$

To gain deeper insights into model behavior, a confusion matrix was constructed based on the XGBoost classifier's predictions on the test set (Figure 4). The matrix reveals that most fault categories were accurately classified, with particularly high precision and recall for the broken rotor bars and no fault classes. Minimal confusion was observed between these classes and other fault types, indicating that their vibrational signatures are sufficiently distinctive.

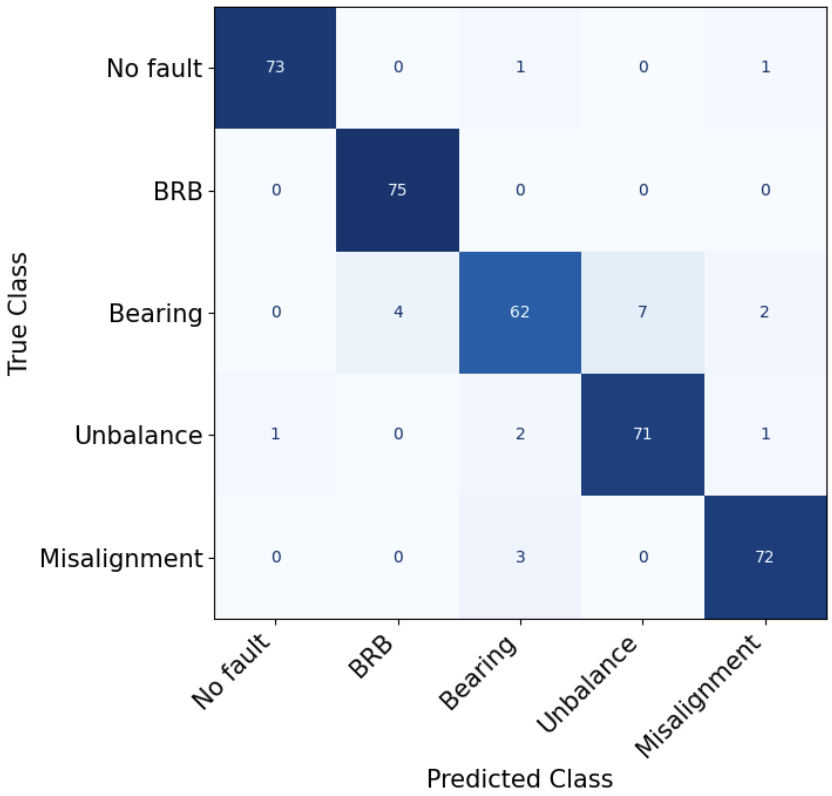


Figure 4. XGBoost confusion matrix for vibration.

However, several misclassifications were noted between classes with overlapping dynamic characteristics. For example, instances of bearing fault were occasionally misclassified as broken rotor bars , and unbalance showed confusion with misalignment , which is expected given the quasi-periodic and amplitude-similar nature of their signal profiles. These confusions are consistent with prior findings in literature [6,60,61], where both fault types exhibit similar statistical trends in low-frequency vibration components.

Overall, the classifier achieved a balanced performance, successfully separating all five classes with minimal cross-class interference. The results confirm that even with a limited set of low-complexity time-domain features, the model can capture meaningful diagnostic patterns and generalize to unseen data.

To complement the confusion matrix analysis and further interpret the inner decision structure of the model, a feature importance evaluation was performed using the trained XGBoost classifier. The resulting importance scores are presented in Figure 5, where the vertical axis quantifies each feature's contribution to the overall model performance, as aggregated across all decision trees.

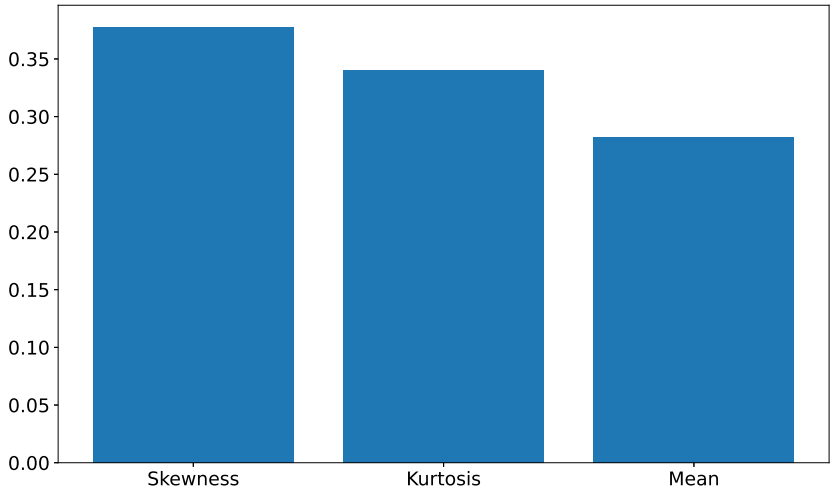


Figure 5. Vibration feature ranking.

Among the three features used, Kurtosis and Skewness was identified as the most influential. This suggests that signal asymmetry, often caused by dynamic imbalance or axis misalignment, plays a critical role in discriminating between fault conditions. Kurtosis followed closely, reinforcing its well-known sensitivity to impulsive events such as those introduced by bearing defects. Mean, while contributing slightly less, remained informative by reflecting signal energy trends that differ across fault types. These results are consistent with the earlier findings of the correlation matrix, which indicated a low statistical redundancy among the selected features. The agreement between statistical independence and the functional importance assigned to the model supports the hypothesis that this compact set of features in the time domain provides both efficient and diagnostically rich input for fault classification.

To enhance the interpretability of the trained classifier, a SHAP (SHapley Additive exPlanations) analysis was conducted. SHAP assigns each feature a contribution score based on its marginal impact on the model output, thereby providing a unified framework for interpreting complex models such as XGBoost. The global explanation based on the mean absolute SHAP values across all test instances is presented in Figure 6. This visualization reveals how much each time-domain feature contributes, on average, to the classification of each fault class.

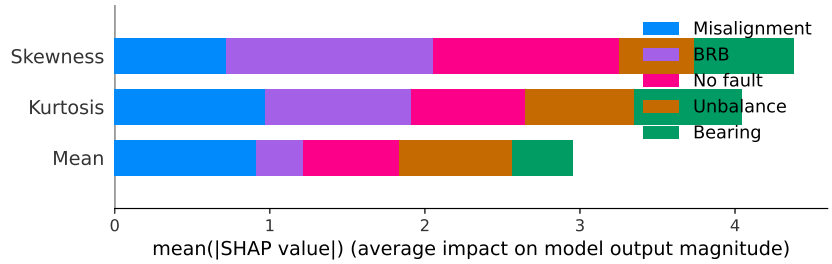


Figure 6. Vibration SHAP value analysis.

Among the three features, Skewness exhibited the highest overall impact, reinforcing its earlier dominance in the feature importance ranking. It played a substantial role in distinguishing no fault, broken rotor bars, and bearing classes, highlighting its relevance in capturing directional asymmetries and waveform shape distortions. Kurtosis followed closely and was most influential in predicting unbalance and no-fault conditions, consistent with its sensitivity to impulsive transients and periodicity shifts. Mean, while contributing

less globally, was particularly active in the classification of bearing and misalignment faults, suggesting its utility in tracking gradual energy shifts.

These results validate the prior selection of statistically independent yet diagnostically complementary features, confirming that each contributes non-redundant information critical for accurate fault differentiation. The SHAP analysis also supports the model's robustness by demonstrating that decision boundaries are informed by interpretable, physically meaningful features, thus enhancing trust in the classifier's outputs.

To evaluate the robustness of the proposed methodology under noisy operating conditions, artificial pink noise was injected into the vibration signals at different signal-to-noise ratio (SNR) levels prior to feature extraction and classification with the XGBoost model. The results indicate that the classifier with statistical features maintained high resilience at moderate noise levels, achieving 93.3% accuracy at 40 dB SNR. However, performance degraded progressively as the noise increased, with accuracy dropping to 80.3% at 30 dB, 54.7% at 20 dB, and only 30.7% at 10 dB. At the most severe condition of 5 dB SNR, classification accuracy fell to 24.3%, approaching random-guessing performance.

In addition to evaluating statistical features, we considered whether PCA could be applied directly to raw time-domain signal windows. This approach eliminates the need for handcrafted feature engineering, instead projecting each window into a low-dimensional latent space using a precomputed projection matrix. The eigenvalue spectrum revealed that the first two principal components captured the majority of the signal variance 49.65% and 45.15%, respectively), with the cumulative explained variance reaching 94.8% after only two components and exceeding 99% by the sixth component. This indicates that the high-dimensional vibration signal can be effectively represented in a very low-dimensional subspace without substantial information loss. The resulting PCA projection matrix, when serialized in fixed-point format, required 41.83 KB of memory, which exceeds the predefined 30 KB limit for edge deployment. However, this overhead can be mitigated by restricting the number of retained principal components, as shown in Table 6, three or four components already preserve more than 97% of the original variance, thereby reducing memory footprint while maintaining diagnostic fidelity. These findings confirm that PCA on raw signals provides a compact representation with strong variance preservation, though the associated storage requirements and computational overhead can be balanced against the constraints of ultra-low-power microcontrollers.

Table 6. PCA of vibration signal.

Component	Eigenvalue	Explained Variance (%)	Cumulative (%)
1	62,758,236.01	49.65	49.65
2	57,062,014.73	45.15	94.80
3	3,619,120.57	2.86	97.66
4	889,888.96	0.70	98.36
5	582,736.74	0.46	98.83
6	557,847.27	0.44	99.27
7	287,418.55	0.23	99.49
8	270,233.25	0.21	99.71
9	129,026.15	0.10	99.81
10	101,984.31	0.08	99.89

To further examine the feasibility of applying PCA-based dimensionality reduction to raw vibration signals on the RP2040, an empirical experiment was conducted using four retained principal components. A window of 254 samples was processed and projected into the reduced subspace. The execution on the microcontroller required 132.23 KB of working memory and 153 ms of computation time, both of which remained within the

operational limits of the device for this configuration. The choice of four components ensured that more than 98% of the original signal variance was preserved, providing a faithful low-dimensional representation suitable for subsequent fault classification tasks.

To further assess the diagnostic value of PCA-derived representations, the raw vibration signals were first projected into a low-dimensional subspace using principal component analysis and subsequently classified with XGBoost (Table 7). The model achieved an overall accuracy of 94.7%, with macro-averaged precision, recall, and F1 score all at 0.95. The PCA-based approach yielded a compact model footprint of 0.37 MB, which was further reduced to 0.15 MB after model distillation, confirming its suitability for deployment on memory-constrained embedded platforms.

Table 7. XGBoost classification report using PCA-transformed raw vibration features.

Class	Precision	Recall	F1 Score	Support
No fault (0)	1.00	1.00	1.00	75
Broken rotor bars (1)	0.83	0.92	0.87	75
Bearing fault (2)	0.91	0.81	0.86	75
Unbalance (3)	1.00	1.00	1.00	75
Misalignment (4)	1.00	1.00	1.00	75
Overall accuracy		0.95 (375 test samples)		
Macro average	0.95	0.95	0.95	375
Weighted average	0.95	0.95	0.95	375

Model size: 0.37 MB; distilled model size: 0.15 MB.

When comparing PCA-derived representations with hand-crafted statistical time-domain features, the results demonstrated only marginal differences in diagnostic accuracy. Specifically, the PCA-based XGBoost classifier achieved an overall accuracy of 94.67%, while the model trained on statistical features (MAV, RMS, Skewness, Kurtosis, and Shape Factor) reached 94.13%. The small numerical gap between the two approaches is well within the expected variability introduced by hyperparameter optimization using Optuna, suggesting that the difference is not statistically significant. In practical terms, this indicates that PCA did not provide a tangible advantage over the direct use of statistical features for vibration-based fault classification. Artificial pink noise was injected into vibration signals at different levels of the SNR and the classification performance was compared to that obtained using handcrafted statistical features. The results demonstrate that PCA projections provide slightly different resilience patterns compared to statistical descriptors. At high noise levels (40 dB), PCA exhibited a lower classification accuracy (90.5%) compared to statistical features (93.3%), suggesting that handcrafted features retained more discriminative information under mild noise. However, as the noise intensity increased, PCA showed a relatively slower degradation. For example, at 20 dB SNR, PCA retained 71.7% accuracy, substantially outperforming the statistical features (54.7%). Similarly, at 10 dB and 5 dB SNR, PCA achieved 55.1% and 43.5%, respectively, whereas the statistical features dropped to 30.7% and 24.3%. This contrast reveals a nuanced trade-off between the two approaches. While handcrafted statistical features are more efficient and perform slightly better under low-to-moderate noise conditions, PCA provides stronger robustness in highly contaminated scenarios by distributing variance across orthogonal components and suppressing random perturbations in individual signal dimensions.

The classification results based on PCA-transformed components extracted from the motor current signal are presented in Figure 7. The confusion matrix indicates that the XGBoost classifier achieved nearly perfect discrimination across most fault classes, with only minor misclassifications observed in the case of bearing faults, which were occasionally

confused with unbalance and misalignment. All other classes, including no fault, broken rotor bars, unbalance, and misalignment, were recognized with complete accuracy.

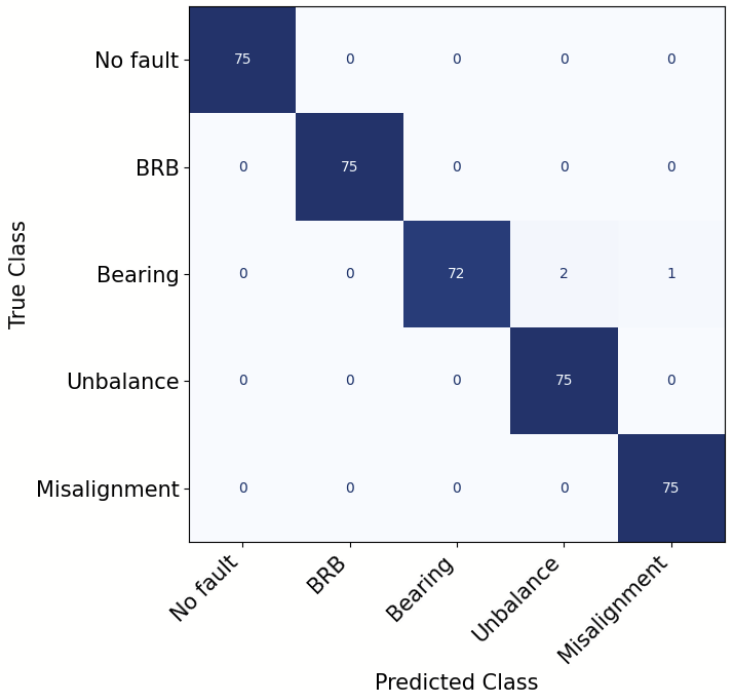


Figure 7. MCSA confusion matrix using PCA components.

4.2. Current Features Validation

To evaluate the effectiveness of time-domain features in current-based fault diagnostics, stator current signals were acquired using non-invasive SCT013-005 current sensors connected to all three phases of a three-phase induction motor. The data acquisition system sampled the signals at a rate of 2 kHz and the resulting measurements were stored in CSV format, with separate files corresponding to distinct fault types.

The dataset consists of five labeled signal segments, each corresponding to one of the following motor conditions: no fault, broken rotor bars (BRB), bearing fault (outer race), mechanical rotor imbalance, and combined misalignment. Each segment contains 60,000 samples per phase, totaling 300,000 samples across all classes.

The signals were loaded into a single data structure for processing and feature extraction. Each row in the dataset corresponds to a timestamped measurement from all three phases—denoted as L1, L2, and L3—alongside the time axis.

The most useful time-domain features for MCSA include energy, RMS [33,62], Kurtosis, Skewness [63], local extrema, and sample extrema. These features provide high resolution, robustness to noise, and cost-effective monitoring, making them ideal for real-time fault detection in induction motors. Additionally, integrating symmetrical components and employing dimensionality reduction techniques further enhances the diagnostic capabilities of MCSA [6].

In this work, four statistical time-domain features were used that are most relevant: Root Mean Square, Skewness, Shape Factor, and Kurtosis. Those features are analyzed in the context of Motor MCSA. These features also have proven effectiveness in capturing waveform energy, asymmetry, impulsiveness, and geometric shape changes, all of which are known to be affected by common motor faults.

To analyze the statistical dependencies among selected time-domain features in current signals, a correlation matrix was computed using measurements from a single phase (L1). This decision was made for several practical and diagnostic reasons. Firstly, due to the

symmetry of the three-phase motor under normal operating conditions, all three phases exhibit similar current signatures in the absence of faults. In many documented cases, faults such as broken rotor bars or eccentricity tend to disturb one or two phases more significantly, but the diagnostic features extracted from each phase remain qualitatively similar [64]. Therefore, analyzing a single phase is sufficient to reveal the statistical properties of the signal for the purpose of feature selection. Secondly, the use of one phase simplifies the processing pipeline and reduces computational load—an essential consideration for embedded applications such as those targeting the RP2040 platform. By computing features on the L1 phase alone, we can preserve diagnostic value while maintaining minimal resource usage.

The features Root Mean Square, Skewness, Shape Factor (SF), and Kurtosis were extracted from the L1 signal using a 100 ms sliding window with no overlap. The resulting correlation matrix is shown in Figure 8. The analysis indicates that all features exhibit weak mutual correlation, with the strongest pairwise coefficient observed between RMS and Skewness ($r = 0.31$). Shape Factor shows mild negative correlation with both RMS and Skewness, suggesting it captures complementary geometric characteristics of the signal. Kurtosis remains statistically independent from all other features, supporting its value as a measure of impulsiveness uncorrelated with energy or symmetry metrics.

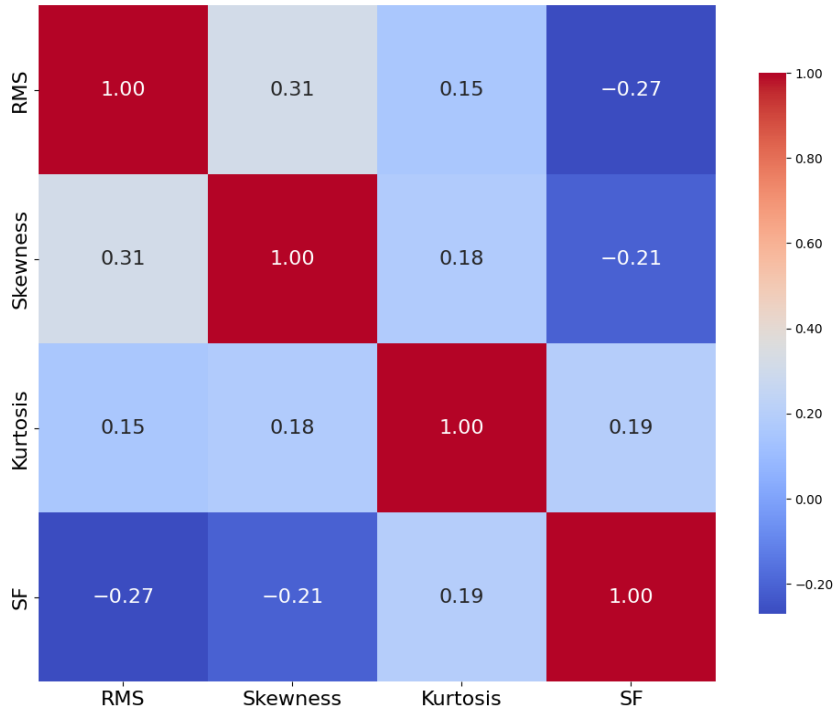


Figure 8. MCSA feature correlation matrix.

In addition to Pearson correlation, a nonparametric Spearman rank correlation analysis was also performed to assess the robustness of the findings with respect to outliers and non-linear monotonic dependencies. The Spearman correlation coefficients among the selected features remained consistently low, with absolute values not exceeding 0.40. This reinforces the conclusion that the features are statistically independent, even under non-Gaussian signal distributions or impulsive conditions typically encountered in motor current signals.

The low pairwise correlation confirms that the selected features capture diverse and non-redundant aspects of the current waveform, supporting their joint use in classification models. The use of only one phase further demonstrates that MCSA-based diagnostics can remain effective even under resource constraints.

For training in the classification model, statistical features were extracted from the L1 phase of each current signal file using a sliding window of 200 samples (corresponding to 100 ms at a sampling rate of 2 kHz). The window was advanced without overlap to ensure non-redundant observations. Within each window, four features were computed: Root Mean Square, Shape Factor, Skewness, and Kurtosis.

Each extracted feature vector was labeled with the corresponding fault class, resulting in a balanced dataset suitable for supervised learning. The full dataset was then stratified and split into training and testing subsets using a 75:25 ratio to ensure consistent class distribution across both sets.

The XGBoost classifier trained on unimodal stator current signals (L1 phase only) achieved a balanced and compact model suitable for embedded deployment. The optimal configuration, identified via Bayesian hyperparameter optimization, included a maximum tree depth of 3, learning rate of 0.205, and 62 boosted estimators.

The model achieved a classification accuracy of 83% on the test set, demonstrating balanced predictive capacity across multiple fault types. As detailed in Table 8, the classifier showed perfect performance on the no fault and misalignment classes, with F1 scores of 1.00 and 0.99, respectively. Performance on broken rotor bars and rotor imbalance remained acceptable (F1: 0.73–0.78), while the bearing fault class exhibited the lowest score (F1 = 0.64), suggesting potential signal overlap with other fault types or insufficient temporal localization.

Table 8. XGBoost classification report using MCSA features.

Class	Precision	Recall	F1 Score	Support
No fault (0)	1.00	1.00	1.00	75
Broken rotor bars (1)	0.76	0.80	0.78	75
Bearing fault (2)	0.65	0.64	0.64	75
Unbalance (3)	0.76	0.71	0.73	75
Misalignment (4)	0.97	1.00	0.99	75
Overall accuracy		0.83 (375 test samples)		
Macro average	0.83	0.83	0.83	375
Weighted average	0.83	0.83	0.83	375

Model size: 0.29 MB; distilled model size: 0.11 MB.

Classifier demonstrates strong potential for practical fault detection using only a single-phase current input. The resulting model is not only accurate but also compact, with a final serialized size of 0.29 MB and a distilled representation of 0.11 MB, well within the storage constraints of RP2040. This confirms the feasibility of deploying diagnostic on embedded hardware.

To provide a fault-specific evaluation of the classifier's performance, a confusion matrix was constructed for the model trained on unimodal MCSA features extracted from the L1 phase. The results, shown in Figure 9, reveal both the strengths and limitations of the classifier across different fault types.

The analysis confirms that the selected features exhibit low pairwise correlations, with absolute values below 0.32. RMS and SF are mildly anticorrelated, suggesting that energy content and waveform sharpness represent distinct signal properties. Skewness and Kurtosis show weak positive correlations with other features, indicating that impulsiveness and asymmetry vary independently of average energy or signal shape.

To gain insights into the relative contribution of each time-domain feature to the classification process, a feature importance analysis was conducted based on the trained XGBoost model. The importance scores were derived from the model's internal split gain statistics and are presented in Figure 10.

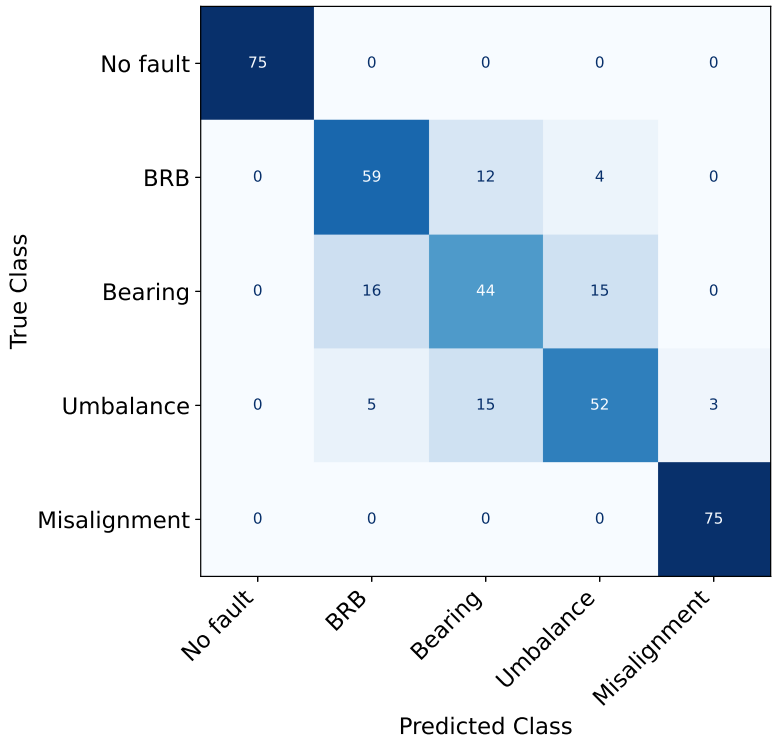


Figure 9. MCSA confusion matrix using statistical features.

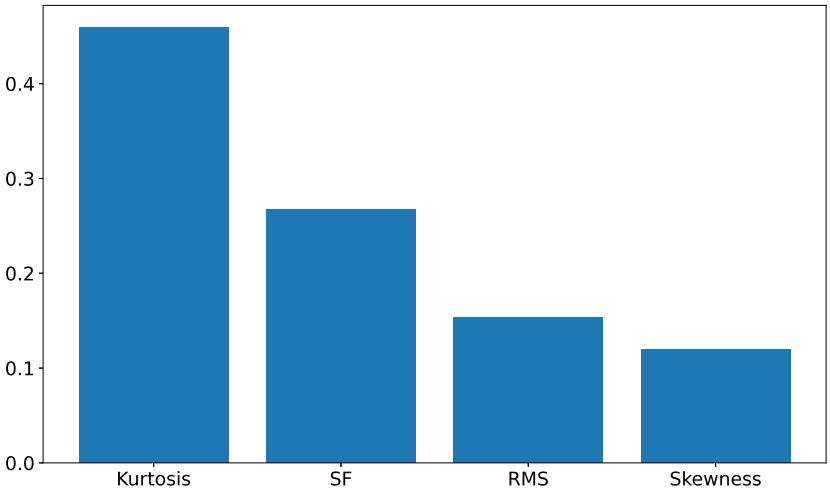


Figure 10. MCSA statistical features ranking.

Among the four characteristics analyzed, the Root Mean Square, Shape Factor, Skewness, and Kurtosis emerged as the most influential, accounting for approximately 46.8% of the total importance. This result confirms the critical role of signal impulsiveness in distinguishing between different fault types, particularly those involving rolling element impacts or rotor asymmetries. The Shape Factor, which quantifies the geometric complexity of the waveform, ranked second (27.1%), highlighting its ability to separate faults that manifest as subtle modulations in current curvature, such as eccentricity or misalignment. Skewness contributed the least (11.4%), indicating that waveform asymmetry, though relevant, is less decisive than impulsiveness or shape for the classification task.

To further interpret the model internal reasoning and understand the class-specific contribution of each feature, SHAP analysis was performed. Figure 11 presents the global bar plot of the mean absolute SHAP values for each feature, aggregated across all classes.

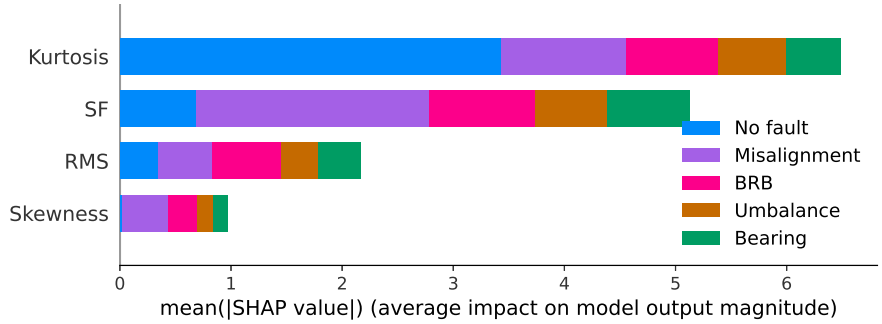


Figure 11. MCSA SHAP value analysis.

The results closely match the ranking produced by the built-in XGBoost feature importance metric, but provide additional granularity by showing how each feature contributes to the prediction of specific fault types. Kurtosis emerged as the most influential feature overall, exhibiting high SHAP contributions across all classes. It was particularly dominant in distinguishing misalignment, BRB, and unbalance, which are known to introduce impulsive and spiky distortions in the stator current waveform. Its role in predicting no-fault conditions was also pronounced, suggesting that a lack of impulsiveness is a strong indicator of healthy operation. Shape Factor (SF) showed substantial influence as well, especially for the misalignment, BRB, and bearing classes. SF captures curvature and amplitude modulation patterns, which are characteristic of mechanical distortions such as rotor imbalance or shaft misalignment. Its diagnostic reach across both electromagnetic and mechanical fault types highlights its versatility. RMS and Skewness had lower SHAP contributions but still demonstrated targeted influence. RMS was moderately relevant for bearing and unbalance classification, consistent with its ability to reflect energy shifts under mechanical stress. Skewness exhibited the lowest overall impact, contributing marginally to class separation, possibly due to the weak asymmetry exhibited in certain faults such as BRB.

Applying PCA to the motor current signals revealed that the first principal component alone accounted for 74.35% of the total variance, while the first three components together explained 83.7% (Table 9). The cumulative variance exceeded 90% with five retained components and reached 95.2% with ten components, indicating that the high-dimensional current waveform could be represented in a low-dimensional subspace with minimal information loss. The serialized PCA projection matrix used 41.83 KB in memory, exceeding the predefined compact threshold of 30 KB. Using the first five principal components reduces projection size to 22.72 KB.

Table 9. PCA of raw current signal.

Component	Eigenvalue	Explained Variance (%)	Cumulative (%)
1	581,258.69	74.35	74.35
2	37,269.41	4.77	79.12
3	35,679.92	4.56	83.68
4	28,089.81	3.59	87.28
5	26,973.26	3.45	90.73
6	22,316.20	2.85	93.58
7	8231.06	1.05	94.63
8	1637.60	0.21	94.84
9	1595.91	0.20	95.05
10	1373.50	0.18	95.22

The XGBoost classifier achieved an overall accuracy of 95.5% (Table 10). The final model size was 0.27 MB, which was reduced to 0.11 MB after distillation, confirming the feasibility of deploying PCA-based current analysis on resource-constrained microcontrollers. Compared to handcrafted statistical features, which yielded only 83.2% classification accuracy, the PCA-based approach demonstrated a clear performance advantage for MCSA.

Table 10. XGBoost classification report using PCA-transformed MCSA features (5 components).

Class	Precision	Recall	F1 Score	Support
No fault (0)	1.00	1.00	1.00	75
Broken rotor bars (1)	0.88	0.96	0.92	75
Bearing fault (2)	0.95	0.83	0.89	75
Unbalance (3)	0.95	1.00	0.97	75
Misalignment (4)	1.00	0.99	0.99	75
Overall accuracy		0.95 (375 test samples)		
Macro average	0.96	0.95	0.95	375
Weighted average	0.96	0.95	0.95	375

Model size: 0.27 MB; distilled model size: 0.11 MB.

Deploying compact models on edge hardware, the performance of the distilled (student) model was compared with its original (teacher) counterpart. The evaluation confirmed that model distillation, based on tree-flattening and structural compression, achieved an identical diagnostic performance while substantially reducing memory footprint. Accuracy and Macro-F1 metrics were preserved to four decimal places, and the probabilistic outputs of both models showed negligible divergence (maximum absolute difference on class probabilities below 1.3×10^{-7}, mean difference 9.7×10^{-9}). The detailed comparison between teacher and student models is summarized in Table 11.

Table 11. Comparison between teacher (original) and student (distilled) XGBoost models.

Metric	Teacher Model	Student Model	Difference (Δ)
Accuracy	0.9547	0.9547	+0.0000
Macro-F1	0.9540	0.9540	+0.0000
Prob. diff (max)	–	–	1.241×10^{-7}
Prob. diff (mean)	–	–	9.727×10^{-9}

4.3. Multimodal Feature Validation

To explore the added diagnostic value of multimodal sensing, current and vibration signals were combined into a unified dataset. Time-synchronized data were collected using three SCT013-005 current sensors and an MPU-6050 triaxial MEMS accelerometer, each sampled at 2 kHz for 10 seconds per condition. The vibration signal was obtained by summing the X, Y, and Z axes to form a single scalar feature representing total vibrational energy at each time step. The corresponding L1-phase current signal was preserved in its raw form. Each pair of vibration and current samples was matched temporally and labeled according to the underlying fault condition. The final dataset comprised three columns—raw current, aggregate vibration, and the ground-truth class label.

Principal component analysis (Figure 12) was performed to evaluate redundancy and dimensional structure in the combined multimodal feature set derived from both current (MCSA) and vibration signals. The first four principal components collectively explained 86.3% of the total variance, with PC1 and PC2 accounting for over 58%. This concentration of information in relatively few components suggests that dimensionality reduction may be feasible without significant information loss.

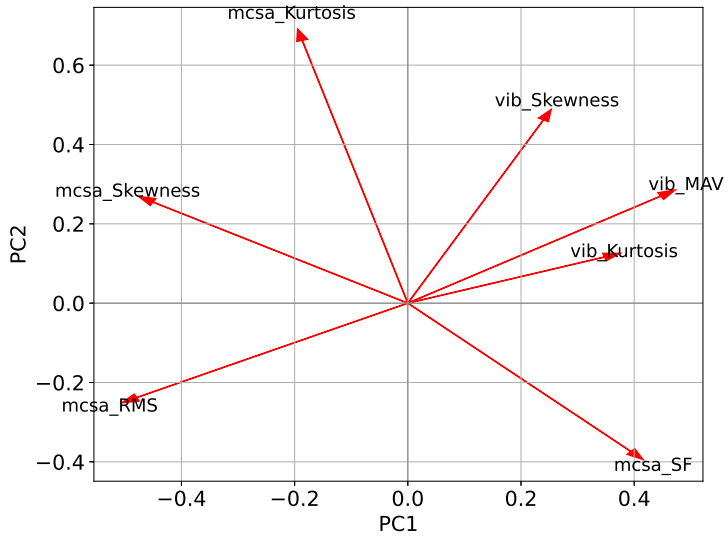


Figure 12. Feature contributions to principal components.

The first principal component was strongly influenced by the Root Mean Square of the stator current and the Mean Absolute Value of the vibration signal. Interestingly, these two features contributed in nearly opposite directions, suggesting that they capture inversely correlated aspects of overall energy content in their respective sensing domains. This relationship probably reflects an underlying physical mechanism, that is, that elevated mechanical vibration often coincides with reduced electrical load and vice versa. Although this finding supports the physical complementarity of the modalities, it also raises the question of redundancy, as both features appear to encode overlapping amplitude-related information.

The second component was dominated by Kurtosis in the current signal and Skewness in the vibration signal. These features capture higher-order shape descriptors related to impulsiveness and asymmetry, respectively. Their joint contribution along a shared axis suggests that waveform non-linearity is an important but multimodal phenomenon, observable in both electrical and mechanical domains. Although both features provide valuable information, their aligned loadings indicate the potential for partial redundancy.

In contrast, Shape Factor from the current signal exhibited a unique loading direction, strongly oriented away from other features in the biplot projection. This suggests it captures a distinct signal characteristic, likely related to waveform curvature, and thus offers complementary information to the overall feature space. Similarly, Skewness in the current signal and Kurtosis in the vibration signal contributed predominantly to the third and fourth components, indicating that they are less dominant but still orthogonal contributors to the signal morphology.

The PCA analysis and the accompanying biplot suggest that the Root Mean Square of current and the Mean Absolute Value of vibration may encode similar amplitude-related information and could therefore be considered functionally redundant. Retaining only one of them, based on computational efficiency or sensor availability, could reduce the set of features without compromising representational capacity. Additionally, partial redundancy is observed between current Kurtosis and vibration Skewness, while the Shape Factor remains a unique and valuable descriptor. These findings support a targeted feature selection strategy in which a reduced but decorrelated subset of features is prioritized for embedded fault classification tasks.

To further reduce model complexity and eliminate redundant features, a feature selection process was integrated into the hyperparameter optimization using the Optuna framework. In this setup, three candidate features—namely, the Root Mean Square of the

current signal, the Mean Absolute Value of the vibration signal, and the Kurtosis of the current waveform—were treated as Boolean decision variables within the optimization space. The objective function simultaneously optimized classifier accuracy and constrained the final model size to remain below 200 kilobytes, in accordance with embedded system memory limitations.

The results presented in Table 12 provide clear insights into the diagnostic value and potential redundancy of selected time-domain features across multimodal sensor data. The highest macro-averaged F1 score of 0.960 was achieved when all three optional features—Root Mean Square of the current, Kurtosis of the current, and Mean Absolute Value of vibration—were included, indicating that a complete feature set yields the most robust classification performance under unconstrained conditions. However, several other feature combinations achieved only slightly lower F1 scores while omitting one of the optional features. For instance, removing either Kurtosis from the current or mean absolute vibration value individually reduced the F1 score by less than 0.045. This marginal performance degradation suggests that these features carry partially overlapping or substitutable information. Specifically, RMS and MAV both characterize signal energy and may exhibit functional redundancy, particularly when both domains respond similarly to fault-induced load variations.

Table 12. Best classification performance per feature subset, based on macro-averaged F1 score.

RMS (MCSA)	Kurtosis (MCSA)	MAV (Vibration)	F1 Score
True	True	True	0.960
True	False	True	0.918
True	True	False	0.917
False	True	True	0.915
True	False	False	0.847
False	True	False	0.836
False	False	True	0.834

Each combination corresponds to a reduced input set used for classification.

Although the principal component analysis revealed moderate to strong correlations between certain features—most notably between the Root Mean Square of the current signal and the Mean Absolute Value of the vibration signal—the empirical results from the joint feature selection experiments tell a more nuanced story. Specifically, while PCA suggested potential redundancy, the removal of amplitude-related or higher-order statistical features led to measurable declines in classification accuracy. The most striking example is the Root Mean Square of the current signal. Despite being highly correlated with vibration energy, models that excluded this feature consistently underperformed compared to those that retained it. Furthermore, this feature is computationally inexpensive to calculate, requiring only a squared mean and square root operation over the window, making it especially attractive for real-time inference on resource-constrained microcontrollers. These findings illustrate a key limitation of relying solely on correlation-based analyses for feature elimination. While statistical correlation indicates similarity in value trends, it does not necessarily imply functional equivalence in a model decision-making process. In this case, the combination of RMS and vibration MAV may encode complementary information that enhances classifier robustness, particularly under ambiguous or transitional fault conditions.

To evaluate the effectiveness of the final optimized feature set, a multimodal classification model was trained using both current-based and vibration-based descriptors. The feature set included Root Mean Square, Shape Factor, Skewness, and Kurtosis from the stator current, along with Mean Absolute Value, Skewness, and Kurtosis from the

vibration signal. Hyperparameter optimization was conducted using Optuna, and the best configuration was selected based on macro-averaged F1 score and model size constraints.

The resulting model achieved a macro-F1 score of 0.96 on the test set, with class-wise F1 scores ranging from 0.92 to 1.00, as summarized in Table 13. Importantly, the distiled XGBoost model occupied only 0.17 MB of memory, well below the 200 KB deployment threshold, confirming its suitability for real-time execution on resource-constrained edge devices such as the RP2040.

Table 13. Classification report for multimodal fault detection (MCSA + Vibration).

Class	Precision	Recall	F1 Score
No Fault	1.00	1.00	1.00
Broken Rotor Bar	0.96	0.95	0.95
Bearing Fault	0.93	0.91	0.92
Unbalance	0.92	0.96	0.94
Misalignment	0.99	0.99	0.99
Macro Average	0.96	0.96	0.96
Overall Accuracy		0.96	

Model size: 0.41 MB; distilled model size: 0.17 MB.

Compared to unimodal baselines, the multimodal configuration demonstrated a clear advantage. A classifier trained solely on current based features achieved a macro-averaged F1 score of 0.83, while the vibration only model reached 0.89. In contrast, the proposed multimodal approach achieved an F1 score of 0.96. This improvement illustrates that the two modalities carry complementary diagnostic information. The performance gain confirms the hypothesis that fault signatures often manifest differently across sensing domains, and that their fusion can mitigate classification ambiguities present in unimodal systems.

This performance gain highlights the complementary nature of electrical and mechanical sensing modalities. While current signals are sensitive to electromagnetic asymmetries, vibration data more effectively captures structural and mechanical faults such as misalignment or unbalance. By combining both modalities, the model benefits from a richer representation of fault signatures, leading to improved classification accuracy and robustness. This confirms the advantage of multimodal fusion in condition monitoring, especially in scenarios where individual modalities may miss subtle or cross-domain fault indicators. Despite the added complexity of feature integration, the final model remained compact and highly deployable, validating the feasibility of low-cost, embedded multimodal fault detection in industrial systems.

The confusion matrix in Figure 13 reveals several important characteristics of the proposed multimodal classifier beyond simple class-wise accuracy. First, the complete absence of false positives in the "No Fault" condition confirms the model's conservative decision boundary when classifying healthy states. This conservative behavior is critical for practical deployment in industrial settings, as it minimizes the risk of false alarms, which are often costly and disruptive.

The matrix demonstrates that multimodal sensing effectively reduces inter-class confusion between fault types with similar statistical profiles. For example, the unbalance and misalignment categories, which exhibited mutual confusion in unimodal models, are now clearly distinguishable. This suggests that the joint feature space learned by the classifier leverages subtle distinctions that are not linearly separable in individual modalities alone. The ability to separate quasiperiodic disturbances (unbalance) from directional misalignments (misalignment) reflects the complementarity of vibration-based asymmetry and current-based waveform curvature descriptors. Only minor misclassifications observed

occurred within fault classes that are known to share overlapping manifestations in their signal morphology, such as bearing faults versus BRB. However, their confusion rates were substantially lower than in unimodal configurations, indicating that multimodal integration not only enhances precision but also improves diagnostic specificity. The results suggest that multimodal fusion introduces a form of regularization in the classifier's internal representations. By combining statistically uncorrelated but semantically aligned features from independent sensors, the model avoids overreliance on any single modality and instead develops a more resilient and generalizable decision logic, essential for real-world edge deployment subject to sensor drift or partial signal degradation.

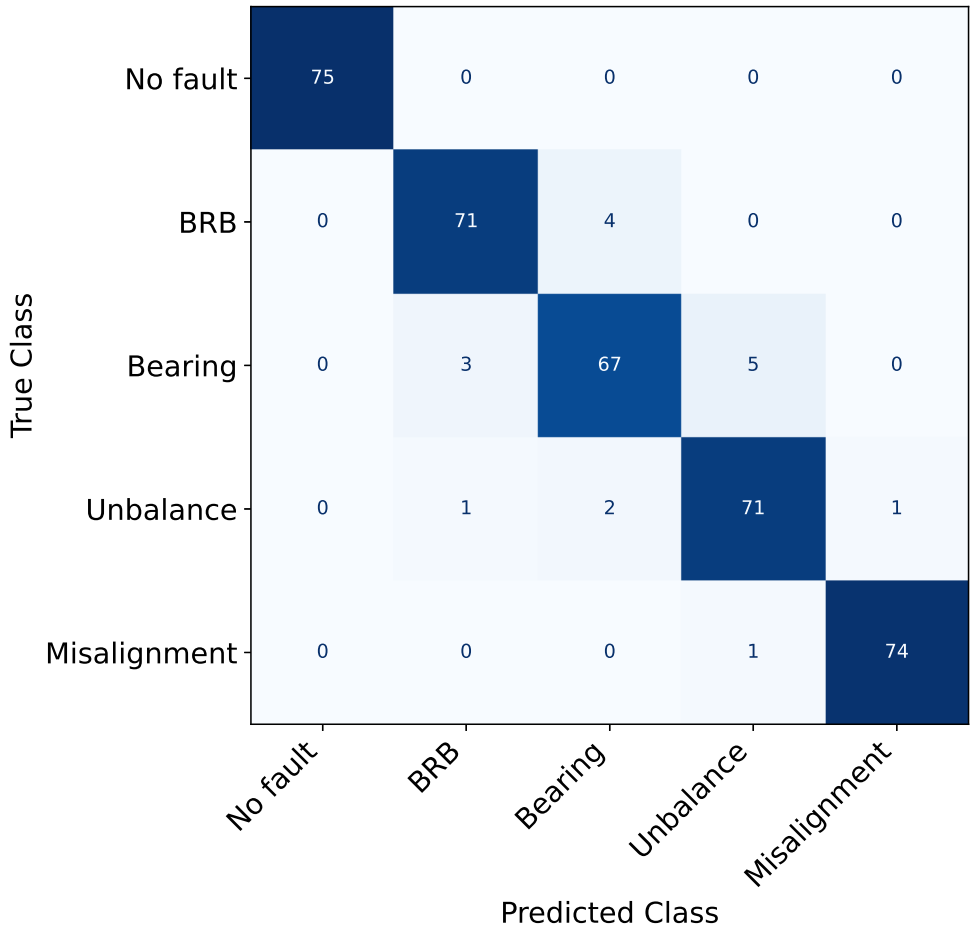


Figure 13. MCSA + vibration confusion matrix.

Figure 14 shows the aggregated feature importance scores derived from the final XGBoost model trained on multimodal features. The highest contribution is attributed to Kurtosis of the stator current signal, indicating its exceptional sensitivity to impulsive transients characteristic of mechanical anomalies such as bearing defects and rotor asymmetries. This is consistent with Kurtosis being a fourth-order moment capable of detecting high-magnitude, low-frequency disturbances which often manifest strongly in current waveforms.

Interestingly, vibration-derived features such as Skewness, Kurtosis, and MAV exhibited uniformly lower importance. This may reflect either redundancy among vibration descriptors or their diminished marginal utility when combined with stronger electromagnetic indicators. However, the relatively balanced importance scores among the vibration features suggest that none were detrimental and all contributed modestly to the classification.

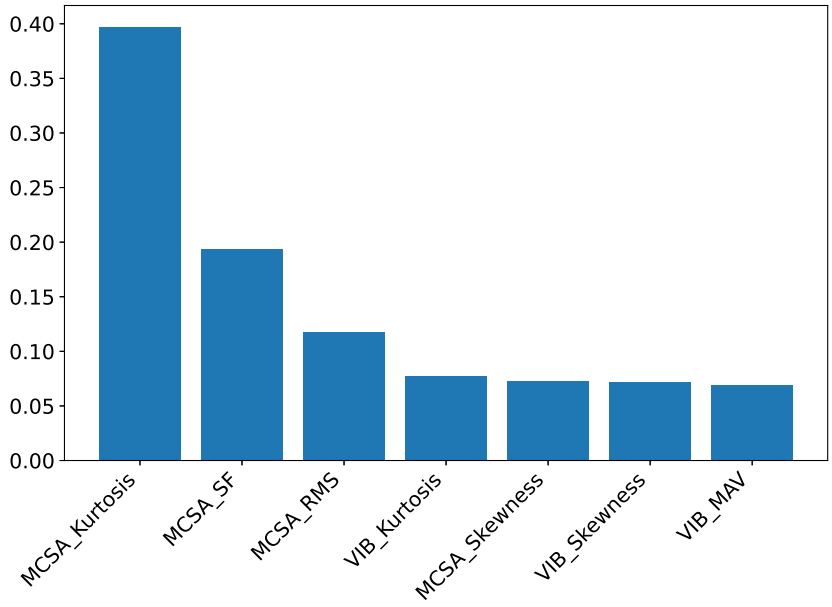


Figure 14. Feature importance of multimodal classification.

4.4. Direct Raw-Signal Classification with XGBoost

When trained directly on raw 200-sample windows from both vibration and motor current signals, the XGBoost classifier demonstrated near-perfect diagnostic performance. The model achieved an overall accuracy of 98.67%, with a macro-averaged F1 score of 0.987 and a logarithmic loss of 0.0513 across five fault categories. These results outperform both PCA-reduced variants and TinyML baselines, indicating that retaining the full raw signal information provides a decisive advantage for gradient-boosted ensembles.

The confusion matrix (Figure 15) illustrates that the classifier correctly identified almost all test windows, with only five misclassifications out of 375 samples. Specifically, BRB instances were confused with bearing outer race faults, one bearing fault was misclassified as BRB, and one misalignment window was mistakenly classified as bearing.

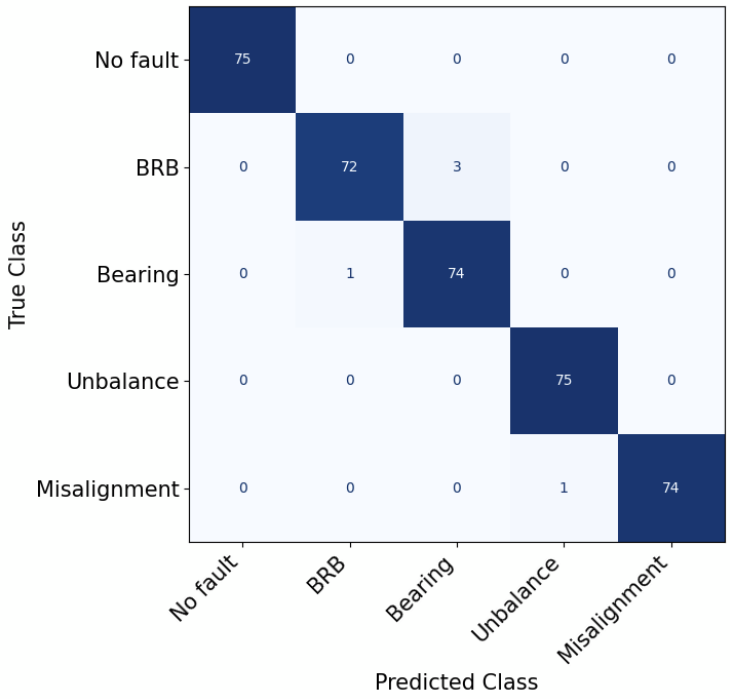


Figure 15. MCSA + vibration confusion matrix raw signal.

Despite the excellent accuracy of the raw signal XGBoost classifier, the absence of dimensionality reduction introduces several drawbacks. First, the input high-dimensionality increases the inference latency compared to PCA-reduced variants. Second, the model footprint grows accordingly: while still deployable (157 KB), it is larger than PCA and statistical feature-based models. Third, raw features are less interpretable, making it more challenging to analyze which signal components are driving classification decisions, whereas PCA components or hand-crafted statistical features provide clearer insights. These limitations indicate that, in practice, a trade-off between accuracy and compactness must be considered, especially when targeting real-time inference on resource-constrained MCUs.

Table 14 summarizes the comparative results before and after feature selection across different modeling strategies.

Table 14. Performance before and after different feature selection strategies (multimodal test data).

Configuration	Accuracy	Model Size (Distilled)
RAW signal	0.987	0.12 MB
Time-domain	0.960	0.17 MB
PCA	0.955	0.11 MB
TinyML (RAW 200D)	0.795	71.44 KB
TinyML (PCA 5D)	0.688	37.88 KB

5. Discussion and Limitations

While the results presented in this study confirm the effectiveness and feasibility of the proposed time-domain feature extraction methodology for embedded predictive maintenance, several limitations should be explicitly recognized.

First, the data set used for model training and validation was collected under laboratory conditions. Such controlled scenarios ensure consistent, high-quality signals but do not fully represent the inherent variability and complexity encountered in real industrial environments. In particular, real-world scenarios involve fluctuating loads, transient events, temperature-induced sensor drift, electromagnetic interference, and combined fault conditions, which were not addressed in this study.

Secondly, this research considered a limited set of artificially induced faults, namely rotor misalignment, faulty bearings, voltage imbalance/single phasing, and broken rotor bars. Although these represent common faults in rotating machinery, many other possible fault conditions and their combinations, such as winding faults, eccentricity variations, and coupled mechanical–electrical anomalies, were not considered, potentially limiting the generalizability of feature extraction and classification models.

In addition, aspects related to sensor placement were not addressed in this study. Optimal placement of current and vibration sensors significantly affects the quality of acquired signals and, consequently, the diagnostic performance. Similarly, the influence of signal conditioning techniques, such as filtering, noise reduction, and signal amplification, was not explicitly analyzed. These preprocessing steps are crucial in real-world conditions, where signal quality can be severely degraded by environmental noise and operational interferences. Future research should explore sensor positioning strategies and systematic signal conditioning techniques to improve the signal integrity and robustness of the diagnostic pipeline under realistic operational conditions.

Although multimodal sensor fusion provided clear advantages in laboratory conditions, the reliability of individual sensors and communication channels in real-world deployments may vary significantly. Sensor degradation, failures, or calibration drift could potentially compromise the quality and consistency of diagnostic results. These practical

aspects of sensor and data reliability are essential areas for further research. Addressing these limitations in future studies, particularly through extensive field testing, exploration of optimal sensor placement, and comprehensive signal conditioning, would significantly strengthen confidence in the viability and robustness of the proposed signal processing methods for embedded predictive maintenance applications.

6. Conclusions

Time-domain statistical features combined with XGBoost achieved an overall accuracy of 94.1% across five fault categories, with balanced macro-F1 of 0.94. PCA applied to vibration signals yielded a very similar performance (94.7% accuracy, macro-F1 = 0.95), suggesting no inherent superiority over handcrafted features in this modality. PCA was highly effective for current signals: classification accuracy improved from 83.2% with statistical features to 95.5% using five principal components, an absolute gain of +12.3 percentage points. This clearly demonstrates that PCA extracts more discriminative structure from MCSA than handcrafted features.

The initial XGBoost models required between 0.27 MB and 0.42 MB, depending on the feature representation. Distillation via tree-flattening reduced them to 0.11–0.19 MB, representing compression ratios of up to 2.5×, while maintaining identical accuracy (differences in accuracy and macro-F1 were within 10^{-4}, with probability differences below 1.3×10^{-7}). This demonstrates that full diagnostic fidelity can be retained in highly compact models. All final models fit comfortably within the 264 KB SRAM limit of the RP2040, leaving operational headroom for signal buffering, preprocessing, and communication tasks. For example, a distilled PCA–MCSA model of 0.11 MB leaves more than 50% RAM capacity available for runtime operations.

In addition, our comparative study with lightweight deep learning baselines (quantized MLP and binary neural networks) confirmed that although TinyML models can operate within tens of kilobytes, their diagnostic accuracy (68–79% for MLP-int8, 31–74% for BNN) remained substantially lower than that of the proposed XGBoost framework. Conversely, training XGBoost directly on raw vibration and current windows yielded near-perfect performance (98.7% accuracy, macro-F1 = 0.987), albeit with larger input dimensionality and model size. This highlights a practical trade-off: while deep learning baselines emphasize compactness, optimized classical ensembles achieve superior diagnostic fidelity on raw signals, and dimensionality reduction remains a viable strategy to balance accuracy with memory efficiency.

The experimental findings demonstrate that fault detection with classification accuracies exceeding 95% can be achieved on microcontrollers with limited computational resources. By constraining the model size to 0.11 MB through distillation, the proposed approach operates within the strict memory budget of the RP2040 (264 KB RAM, 2 MB Flash) while preserving diagnostic fidelity. These results substantiate the feasibility of deploying predictive maintenance pipelines on ultra-low-cost embedded hardware, thereby extending the applicability of intelligent condition monitoring to resource-constrained industrial environments.

7. Future Work

Future research should expand the present study in several directions to further strengthen the applicability of resource-constrained predictive maintenance systems. First, the distillation of diverse state-of-the-art machine learning models into compact gradient-boosted ensembles should be systematically explored, thereby quantifying the trade-offs between diagnostic fidelity, memory footprint, and inference latency. Second, the applicability of alternative dimensionality reduction techniques beyond PCA—such as linear discrimi-

nant analysis, autoencoder-based embeddings, or wavelet-based compression—should be evaluated within the proposed framework to assess their diagnostic value under severe resource constraints. Third, energy consumption remains a critical but underexplored factor in embedded diagnostics. Future work should investigate the relationship between power usage and the choice of feature extraction, dimensionality reduction, and classification methods. Fourth, another important research avenue concerns compound faults, where multiple failure modes co-occur and interact non-linearly, obscuring individual fault signatures and creating diagnostic ambiguity. Addressing compound faults on ultra-constrained MCUs remains largely unexplored, as most existing methods are designed for single-fault scenarios. Extending the proposed framework to systematically simulate and classify compound fault conditions will enable the assessment of how modular combinations of feature extraction, dimensionality reduction, and compact classifiers can disentangle overlapping patterns while respecting strict resource budgets. Finally, the optimization of feature computation routines specifically tailored to the RP2040 microcontroller, including integer-based implementations and parallel execution strategies, could further enhance efficiency and enable deployment in ultra-low-power industrial scenarios. Collectively, these directions offer a pathway toward more robust, energy-efficient, and scalable fault detection solutions for cost-sensitive environments.

Author Contributions: Methodology, R.Ū.; Validation, A.S.; Investigation, R.Ū.; Writing – original draft, R.Ū.; Supervision, A.S. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are not publicly available due to privacy restrictions but are available on request from the corresponding author.

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

References

1. Günal, S.; Gökhan Ece, D.; Nezih Gerek, Ö. Induction Machine Condition Monitoring Using Notch-Filtered Motor Current. *Mech. Syst. Signal Process.* **2009**, *23*, 2658–2670. [CrossRef]
2. Bellini, A.; Yazidi, A.; Filippetti, F.; Rossi, C.; Capolino, G.A. High Frequency Resolution Techniques for Rotor Fault Detection of Induction Machines. *IEEE Trans. Ind. Electron.* **2008**, *55*, 4200–4209. [CrossRef]
3. Divya, D.; Marath, B.; Santosh Kumar, M. Review of Fault Detection Techniques for Predictive Maintenance. *J. Qual. Maint. Eng.* **2023**, *29*, 420–441. [CrossRef]
4. Karamizadeh, S.; Abdullah, S.M.; Manaf, A.A.; Zamani, M.; Hooman, A. An Overview of Principal Component Analysis. *J. Signal Inf. Process.* **2013**, *4*, 173–175. [CrossRef]
5. Neupane, D.; Bouadjenek, M.R.; Dazeley, R.; Aryal, S. Data-Driven Machinery Fault Diagnosis: A Comprehensive Review. *Neurocomputing* **2025**, *627*, 129588. [CrossRef]
6. Nandi, S.; Toliyat, H.; Li, X. Condition Monitoring and Fault Diagnosis of Electrical Motors—A Review. *IEEE Trans. Energy Convers.* **2005**, *20*, 719–729. [CrossRef]
7. Randall, R.B.; Antoni, J. Rolling Element Bearing Diagnostics—A Tutorial. *Mech. Syst. Signal Process.* **2011**, *25*, 485–520. [CrossRef]
8. Donolo, P.; Bossio, G.; De Angelo, C.; García, G.; Donolo, M. Voltage Unbalance and Harmonic Distortion Effects on Induction Motor Power, Torque and Vibrations. *Electr. Power Syst. Res.* **2016**, *140*, 866–873. [CrossRef]
9. Verma, A.K.; Sarangi, S.; Kolekar, M.H. Misalignment Fault Detection in Induction Motor Using Rotor Shaft Vibration and Stator Current Signature Analysis. *Int. J. Mechatron. Manuf. Syst.* **2013**, *6*, 422. [CrossRef]
10. Wang, Z.; Li, S.; Xuan, J.; Shi, T. Biologically Inspired Compound Defect Detection Using a Spiking Neural Network with Continuous Time–Frequency Gradients. *Adv. Eng. Bioinform.* **2025**, *65*, 103132. [CrossRef]
11. Wang, Z.; Xuan, J.; Shi, T. An Autonomous Recognition Framework Based on Reinforced Adversarial Open Set Algorithm for Compound Fault of Mechanical Equipment. *Mech. Syst. Signal Process.* **2024**, *219*, 111596. [CrossRef]

12. Peng, C.; Shangguan, W.; Wang, Z.; Peng, J.; Chai, L.; Xing, Y.; Cai, B. Reliability Assessment of Urban Rail Transit Vehicle On-Board Controller with Multi-Component Failure Dependence Based on R-vine-copula. *Reliab. Eng. Syst. Saf.* **2025**, *257*, 110795. [CrossRef]
13. Tang, H.; Lu, S.; Qian, G.; Ding, J.; Liu, Y.; Wang, Q. IoT-Based Signal Enhancement and Compression Method for Efficient Motor Bearing Fault Diagnosis. *IEEE Sens. J.* **2021**, *21*, 1820–1828. [CrossRef]
14. Han, H.; Siebert, J. TinyML: A Systematic Review and Synthesis of Existing Research. In Proceedings of the 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Jeju Island, Republic of Korea, 21–24 February 2022; pp. 269–274. [CrossRef]
15. Hao, M.; Wang, X.; Li, P.; He, F.; Yang, L.; Xiao, M.; Bian, X.; Zhang, Z. Motor Fault Diagnosis System Based on Single Chip Microcomputer and Artificial Intelligence. In Proceedings of the 2021 IEEE International Conference on Mechatronics and Automation (ICMA), Takamatsu, Japan, 8–11 August 2021; pp. 1413–1418. [CrossRef]
16. Quan, Y.; Lin, F.; Luo, D.; Xue, W. Design of Embedded Motor Diagnostic System Based on Vibration Signal Analysis and Machine Learning. In Proceedings of the 2024 6th International Conference on Telecommunications and Communication Engineering, Chengdu, China, 22–24 November 2024; pp. 96–100. [CrossRef]
17. Harrabi, M.; Hamdi, A.; Ouni, B.; Bel Hadj Tahar, J. Real-time temperature anomaly detection in vaccine refrigeration systems using deep learning on a resource-constrained microcontroller. *Front. Artif. Intell.* **2024**, *7*, 1429602. . [CrossRef]
18. Kim, S.; Lee, S.; Lee, J.; Kim, M.; Kim, S.J.; Yoon, H.; Youn, B.D. Fault-Relevance-Based, Multi-Sensor Information Integration Framework for Fault Diagnosis of Rotating Machineries. *Mech. Syst. Signal Process.* **2025**, *232*, 112742. [CrossRef]
19. Yu, J. Mechatronics Fault Prediction and Diagnosis Based on Multi Sensor Information Fusion. *J. Phys. Conf. Ser.* **2021**, *1982*, 012100. [CrossRef]
20. Tao, Z.; Xia, P.; Huang, Y.; Xiao, D.; Wuang, Y.; Zhong, Z.; Liu, C. Induction Motor Fault Diagnosis Based on Multi-Sensor Fusion Under High Noise and Sensor Failure Condition. In Proceedings of the 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), Nanjing, China, 15–17 October 2021; pp. 1–8. [CrossRef]
21. Gana, M.; Achour, H.; Belaid, K.; Chelli, Z.; Laghrouche, M.; Chaouchi, A. Non-Invasive Intelligent Monitoring System for Fault Detection in Induction Motor Based on Lead-Free-Piezoelectric Sensor Using ANN. *Meas. Sci. Technol.* **2022**, *33*, 065105. [CrossRef]
22. Amla, L.J.; R, V.; B, S.; P, B.R.; K, K. IoT-based Fault Detection in 3-Phase Poleline Networks Using Wi-Fi and Flutter Application. In Proceedings of the 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 14–16 May 2025; pp. 361–366. [CrossRef]
23. Marković, N.; Medojević, M. Hybrid Condition Monitoring for IIoT Applications. In Proceedings of the 2025 10th International Conference on Smart and Sustainable Technologies (SpliTech), Bol and Split, Croatia, 16–20 June 2025; pp. 1–5. [CrossRef]
24. Wang, Y.; Yu, Z.; Wu, J.; Wang, C.; Zhou, Q.; Hu, J. Adaptive Knowledge Distillation-Based Lightweight Intelligent Fault Diagnosis Framework in IoT Edge Computing. *IEEE Int. Things J.* **2024**, *11*, 23156–23169. [CrossRef]
25. Chaoraingern, J.; Numsomran, A. Embedded Sensor Data Fusion and TinyML for Real-Time Remaining Useful Life Estimation of UAV Li Polymer Batteries. *Sensors* **2025**, *25*, 3810. [CrossRef]
26. Wang, W.; Ullah, A.; Li, L.; Wang, M. Optimized Binary Neural Networks for Road Anomaly Detection: A TinyML Approach on Edge Devices. *Comput. Mater. Contin.* **2024**, *80*, 527–546. [CrossRef]
27. Zamani, A.R.; Petri, I.; Diaz-Montes, J.; Rana, O.; Parashar, M. Edge-Supported Approximate Analysis for Long Running Computations. In Proceedings of the 2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud), Prague, Czech Republic, 21–23 August 2017; pp. 321–328. [CrossRef]
28. Castanyer, R.C.; Martínez-Fernández, S.; Franch, X. Which Design Decisions in AI-enabled Mobile Applications Contribute to Greener AI? *Empir. Softw. Eng.* **2024**, *29*, 2. [CrossRef]
29. Cerutti, G.; Prasad, R.; Brutti, A.; Farella, E. Neural Network Distillation on IoT Platforms for Sound Event Detection. In Proceedings of the 20th Annual Conference of the International Speech Communication Association INTERSPEECH 2019, Graz, Austria, 15–19 September 2019; pp. 3609–3613. [CrossRef]
30. Xu, M.; Gu, Y.; Poslad, S.; Xue, S. Optimized Computation Combining Classification and Detection Networks with Distillation. In Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–8. [CrossRef]
31. Gao, Z.; Reviriego, P.; Li, X.; Maestro, J.; Zhao, M.; Wang, J. A Fault Tolerant Implementation of the Goertzel Algorithm. *Microelectron. Reliab.* **2014**, *54*, 335–337. [CrossRef]
32. Zaparoli, I.O.; Júnior, A.M.G.; Rossa, A.J.; Soares, I.F.; Kuhnen, G.; Paula, H.D. Electrical Current Analysis for Non-Invasive Early Inter-Turn Short-Circuit Detection in FOC Driven Induction Motors. *Meas. Sci. Technol.* **2025**, *36*, 046117. [CrossRef]
33. Zhao, X.; Wang, P. A Deployable Edge Computing Solution for Machine Condition Monitoring. In Proceedings of the 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Glasgow, UK, 20–23 May 2024; pp. 1–6. [CrossRef]

34. Dhomad, T.A.; Jaber, A.A. Bearing Fault Diagnosis Using Motor Current Signature Analysis and the Artificial Neural Network. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2020**, *10*, 70–79. [CrossRef]
35. Chikkam, S.; Singh, S. Condition Monitoring and Fault Diagnosis of Induction Motor Using DWT and ANN. *Arab. J. Sci. Eng.* **2023**, *48*, 6237–6252. [CrossRef]
36. Wang, Y.; Wang, P. Explainable Machine Learning for Motor Fault Diagnosis. In Proceedings of the 2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Kuala Lumpur, Malaysia, 21 May 2023; pp. 1–6. [CrossRef]
37. Mousavi, S.; Bayram, D.; Seker, S. Current Data Fusion Through Kalman Filtering for Fault Detection and Sensor Validation of an Electric Motor. In Proceedings of the 2019 International Aegean Conference on Electrical Machines and Power Electronics (ACEMP) & 2019 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), Istanbul, Turkey, 2–4 September 2019; pp. 155–160. [CrossRef]
38. Süpürtülü, M.; Hatipoğlu, A.; Yılmaz, E. An Analytical Benchmark of Feature Selection Techniques for Industrial Fault Classification Leveraging Time-Domain Features. *Appl. Sci.* **2025**, *15*, 1457. [CrossRef]
39. Prieto, M.D.; Cirrincione, G.; Espinosa, A.G.; Ortega, J.A.; Henao, H. Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks. *IEEE Trans. Ind. Electron.* **2013**, *60*, 3398–3407. [CrossRef]
40. Motahari-Nezhad, M.; Jafari, S.M. Bearing Remaining Useful Life Prediction under Starved Lubricating Condition Using Time Domain Acoustic Emission Signal Processing. *Expert Syst. Appl.* **2021**, *168*, 114391. [CrossRef]
41. Sánchez, R.V.; Macancela, J.C.; Ortega, L.R.; Cabrera, D.; García Márquez, F.P.; Cerrada, M. Evaluation of Hand-Crafted Feature Extraction for Fault Diagnosis in Rotating Machinery: A Survey. *Sensors* **2024**, *24*, 5400. [CrossRef]
42. Kundur, D. A Generalized Design Framework for IIR Digital Multiple Notch Filters. *Eurasip J. Adv. Signal Process.* **2015**, *2015*, 26. [CrossRef]
43. Sun, H.; Yuan, S.; Luo, Y.; Guo, Y.; Yin, J. Unsteady Characteristics Analysis of Centrifugal Pump Operation Based on Motor Stator Current. *Proc. Inst. Mech. Eng. Part J. Power Energy* **2017**, *231*, 689–705. [CrossRef]
44. Gunal, S.; Gokhan Ece, D.; Gerek, O.N. Induction motor fault diagnosis via current analysis on time domain. In Proceedings of the 2009 IEEE 17th Signal Processing and Communications Applications Conference, Antalya, Turkey, 9–11 April 2009; pp. 488–491. [CrossRef]
45. Nayana, B.R.; Geethanjali, P. Analysis of Statistical Time-Domain Features Effectiveness in Identification of Bearing Faults From Vibration Signal. *IEEE Sens. J.* **2017**, *17*, 5618–5625. [CrossRef]
46. Pradesh, U.; Yadav, O.P.; Pahuja, G.L. Bearing Health Assessment Using Time Domain Analysis of Vibration Signal. *Int. J. Image, Graph. Signal Process.* **2020**, *12*, 27–40. [CrossRef]
47. Ding, C.; Peng, H. Minimum Redundancy Feature Selection from Microarray Gene Expression Data. CSB2003; In Proceedings of the 2003 IEEE Bioinformatics Conference, Stanford, CA, USA, 11–14 August 2003; pp. 523–528. [CrossRef]
48. Han, Y.; Yuan, J.; Luo, Y.; Zou, J.; Qin, X. Detection and Evaluation of Load Oscillations in Induction Motors Based on MCSA. *Proc. Inst. Mech. Eng. Part J. Process. Mech. Eng.* **2023** , *239*, 244–257. [CrossRef]
49. Ali, M.Z.; Shabbir, M.N.S.K.; Liang, X.; Zhang, Y.; Hu, T. Machine Learning-Based Fault Diagnosis for Single- and Multi-Faults in Induction Motors Using Measured Stator Currents and Vibration Signals. *IEEE Trans. Ind. Appl.* **2019**, *55*, 2378–2391. [CrossRef]
50. Rosli, N.S.B.; Ibrahim, R.B.; Ismail, I. Application of Principle Component Analysis in Resolving Influential Factor Subject to Industrial Motor Failure. In Proceedings of the 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation (UKSim), Cambridge, UK, 27–29 March 2018; pp. 21–25. [CrossRef]
51. Wu, L.; Su, S.; Wen, C. Multiple Fault Diagnosis Methods Based on Multilevel Multi-Granularity PCA. In Proceedings of the 2018 International Conference on Control, Automation and Information Sciences (ICCAIS), Hangzhou, China, 24–27 October 2018; pp. 566–570. [CrossRef]
52. Rahmat, F.; Zulkafli, Z.; Ishak, A.J.; Abdul Rahman, R.Z.; Stercke, S.D.; Buytaert, W.; Tahir, W.; Ab Rahman, J.; Ibrahim, S.; Ismail, M. Supervised Feature Selection Using Principal Component Analysis. *Knowl. Inf. Syst.* **2024**, *66*, 1955–1995. [CrossRef]
53. Atif, M.; Khan, F.; Azmat, S.; Bibi, A. AI-Based Speed Adaptive Fault Diagnosis in Induction Motors Using Vibration Signal. In Proceedings of the 2024 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 9–10 December 2024; pp. 1–6. [CrossRef]
54. Zhong, J.; Bao, X.; Xiao, L.; Zheng, Y. A Physical-Assisted Multi-Source Signal Rolling Bearing Fault Diagnosis Method. In Proceedings of the 2023 Global Reliability and Prognostics and Health Management Conference (PHM-Hangzhou), Hangzhou, China, 12–15 October 2023; pp. 1–6. [CrossRef]
55. Feliciani Merizio, I.; Chavarette, F.R.; Fuzaro De Almeida, E. Machine Learning Techniques for Fault Detection in Rotating Mechanical Systems. *Eng. Comput.* **2025**, *42*, 1316–1334. [CrossRef]
56. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review. *IEEE Access* **2020**, *8*, 29857–29881. [CrossRef]

57. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 3–7 August 2019.
58. Kumar, S.; Mukherjee, D.; Guchhait, P.K.; Banerjee, R.; Srivastava, A.K.; Vishwakarma, D.N.; Saket, R.K. A Comprehensive Review of Condition Based Prognostic Maintenance (CBPM) for Induction Motor. *IEEE Access* **2019**, *7*, 90690–90704. [CrossRef]
59. Ren, L.; Li, X.; Ma, H.; Zhang, G.; Huang, S.; Chen, K.; Wang, X.; Yue, W. Lightweight Intelligent Fault Diagnosis Method Based on a Multi-Stage Pruning Distillation Interleaving Network. *Adv. Mech. Eng.* **2024**, *16*, 16878132241281470. [CrossRef]
60. Yen, G.; Lin, K.C. Wavelet Packet Feature Extraction for Vibration Monitoring. *IEEE Trans. Ind. Electron.* **2000**, *47*, 650–667. [CrossRef]
61. Liu, X.; Hong, J.; Zhao, K.; Sun, B.; Zhang, W.; Jiang, J. Vibration Analysis for Fault Diagnosis in Induction Motors Using One-Dimensional Dilated Convolutional Neural Networks. *Machines* **2023**, *11*, 1061. [CrossRef]
62. Obaid, R.; Habetler, T.; Gritter, D. A Simplified Technique for Detecting Mechanical Faults Using Stator Current in Small Induction Motors. In Proceedings of the Conference Record of the 2000 IEEE Industry Applications Conference. Thirty-Fifth IAS Annual Meeting and World Conference on Industrial Applications of Electrical Energy (Cat. No.00CH37129), Rome, Italy, 8–12 October 2000; Volume 1, pp. 479–483. [CrossRef]
63. Matic, D.; Kulic, F. SVM Broken Bar Detection Based on Analysis of Phase Current. In Proceedings of the 2012 15th International Power Electronics and Motion Control Conference (EPE/PEMC), Novi Sad, Serbia, 4–6 September 2012; pp. LS4c.2–1–LS4c.2–4. [CrossRef]
64. El Hachemi Benbouzid, M. A Review of Induction Motors Signature Analysis as a Medium for Faults Detection. *IEEE Trans. Ind. Electron.* **2000**, *47*, 984–993. [CrossRef]

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