

Physics-Informed Machine Learning for Automated Fault Diagnosis in Hydrodynamic Bearings: A Comprehensive Synthetic Data Approach with Ensemble Classification

Technical Report

Department of Mechanical Engineering

Submitted: November 05, 2025

Abstract

This paper introduces an end-to-end machine learning model to perform automated fault detection in hydrodynamic bearings (Palier Fluide Dynamique) and fulfill the highly important industrial problem of predictive maintenance using a unified methodology of physics-based synthetic data generation and ensemble classification. The methodology incorporates 11 separate operational states with 8 single fault conditions, three composite fault conditions and one healthy baseline condition. The physics-based data generator generates 1,430 realistic vibration waveforms with seven independent sources of noise (measurement, electromagnetic interference, pink noise, drift, quantization, sensor drift and impulse), multi-severity fault development (incipient through severe) and operating conditions computed following the Sommerfeld number theory. Based on these signals, 36 time-frequency-wavelet domain features are obtained; 15 of them are determined using minimum redundancy maximum relevance (MRMR) criterion after split in order to avoid data leakage. Three Bayesian hyper-parameter optimization supervised learning algorithms (Support Vector Machine, Random Forest, Neural Network) are trained through 50 iterations. The Random Forest classifier provides the best results at 95.81 percent validation accuracy, 95.33 percent test accuracy, which is balanced per class (macro F1-score: 95.37 percent, mean AUC: 0.9970). Systematic robustness analysis indicates that the system has superior temporal drift tolerance (3.27 percent degradation), intermediate missing feature resistance (6.54PER cent degradation), but is susceptible to sensor noise (16.82 percent degradation). The most dominant difficulty in the error analysis is the mixed fault classification, and 90 percent of misclassifications include the compound failure modes. The full pipeline takes less than 15 minutes, and the generated inference function is production-ready and can be deployed to industrial condition monitoring systems in real-time.

Keywords: Hydrodynamic bearing, fault diagnosis, synthetic data generation, Random Forest, ensemble learning, predictive maintenance, vibration analysis, condition monitoring

Contents

1 General Introduction	4
1.1 Industrial Context and Motivation	4
1.2 Evolution of Fault Diagnosis Methodologies	4
1.3 Hydrodynamic Bearing Fault Phenomenology	5
1.4 Machine Learning Paradigms for Condition Monitoring	5
1.5 Synthetic Data Generation: Motivation and Challenges	6
1.6 Research Objectives and Contributions	6
2 Problem Statement	7
2.1 Technical Formulation	7
2.2 Operational Constraints and Requirements	8
2.3 Challenges Addressed	8
2.4 Success Criteria	9
3 Research Objectives	9
3.1 Overall Objective	9
3.2 Specific Objectives	9
4 Literature Review	10
4.1 Traditional Vibration-Based Fault Diagnosis	10
4.2 Machine Learning for Condition Monitoring	11
4.3 Synthetic Data Generation for Diagnostics	11
4.4 Multi-Fault and Mixed-Fault Classification	12
4.5 Robustness and Adversarial Evaluation	13
4.6 Gap Analysis and Positioning	13
5 Originality of the Work	14
5.1 Core Methodological Innovations	14
5.2 Comparative Differentiation	15
5.3 Expected Impact and Significance	15
6 Methodology	16
6.1 Problem Domain & Industrial Context	16
6.2 System Classification & ML Paradigm	17
6.3 Fault Taxonomy: 11-Class Classification Problem	17
7 Data Generation System: Physics-Informed Synthetic Signals	17
7.1 Architectural Design Philosophy	18
7.2 Signal Generation Parameters	18
7.3 Physics-Based Fault Modeling Framework	18
7.3.1 Example 1: Misalignment (Désalignement)	18
7.3.2 Example 2: Oil Whirl	19
7.3.3 Example 3: Cavitation	19
7.4 Multi-Severity Fault Progression	19
7.5 Operating Condition Variations	20
7.6 Comprehensive Noise Modeling (7 Independent Sources)	20
7.7 Data Augmentation Strategy	20

8 Feature Engineering: Signal-to-Intelligence Transformation	21
8.1 Multi-Domain Feature Extraction Philosophy	21
8.2 Base Feature Set (36 Features - Fast Extraction)	21
8.2.1 Time-Domain Features (7 features)	21
8.2.2 Envelope Analysis Features (4 features)	22
8.2.3 Frequency-Domain Features (9 features)	22
8.2.4 Harmonic Ratio Features (3 features)	22
8.2.5 Wavelet & Cepstral Features (7 features)	22
8.2.6 Bispectrum & Nonlinearity Features (6 features)	22
8.3 Advanced Feature Set (Additional 16 Features - Optional)	23
8.4 Feature Selection Strategy	23
9 Machine Learning Model Portfolio	23
9.1 Ensemble Learning Strategy	23
9.2 Support Vector Machine (SVM) Configuration	23
9.3 Random Forest Configuration	24
9.4 Neural Network Architecture	24
9.5 Hyperparameter Optimization Protocol	25
10 Validation & Robustness Evaluation Framework	25
10.1 Standard Evaluation Metrics	25
10.2 Advanced Robustness Testing	25
10.2.1 Test 1: Sensor Noise Tolerance	25
10.2.2 Test 2: Missing Feature Robustness	26
10.2.3 Test 3: Temporal Drift Simulation	26
10.3 Confusion Matrix Analysis	26
11 Experimental Results and Discussion	26
11.1 Dataset Characteristics and Generation	26
11.2 Feature Extraction and Selection	27
11.3 Model Training and Selection	29
11.4 Classification Performance: Test Set Results	30
11.5 Confusion Matrix Analysis	31
11.6 ROC Analysis and Discrimination Quality	33
11.7 Robustness Evaluation Under Operational Degradation	34
11.7.1 Test 1: Sensor Noise Tolerance	34
11.7.2 Test 2: Missing Feature Robustness	34
11.7.3 Test 3: Temporal Drift Resilience	35
11.8 Misclassification Analysis and Failure Mode Investigation	36
11.9 Discussion: Performance in Context	37
12 Project Timeline and Task Allocation	38
13 Conclusion and Future Directions	40
13.1 Summary of Achievements	40
13.2 Key Contributions to the Field	40
13.3 Limitations and Constraints	41
13.4 Recommendations for Industrial Deployment	41
13.5 Future Research Directions	42
13.6 Closing Remarks	44
14 References	44

1 General Introduction

1.1 Industrial Context and Motivation

Hydrodynamic bearings represent a critical component class in high performance rotating machines, finding wide-spread application in turbines, compressors, high-speed motors and precision industrial equipment where rolling element bearings are not enough. Unlike their rolling element counterparts, hydrodynamic bearings, also termed fluid film or journal bearings, operate on the principle of separating loaded surfaces through pressurized lubricant films generated by relative motion. This fundamental design philosophy offers distinct advantages including unlimited fatigue life under proper operation, minimal vibration transmission, tolerance to transient overload and accommodation of thermal expansion. Latest power-generation facilities, aerospace propulsion systems and petro=chemical refineries depends heavily on these components to sustain continuous operations at very high speeds and loads.

The economic stakes surrounding bearing reliability are enormous. Out of schedule down-times in industrial facilities leads to losses ranging from \$5,000 to \$50,000 per hour depending on the scale. Apart from direct financial impact, catastrophic bearing failures creates real safety hazards. Such as potential fire risk from lubricant ignition, equipment destruction from rotor-stator contact, and environmental contamination from oil leakage. The U.S. Department of Energy states that failure related to bearing makes up for roughly 40-50% of rotating machinery breakdowns. This underscores the criticality of effective condition monitoring strategies.

Traditional maintenance approaches fall into two categories: reactive (run-to-failure) and preventive (time-based replacement). The former accepts downtime costs and collateral damage, while the latter discards functional components prematurely. Predictive maintenance paradigms using condition monitoring offer a third way, intervening based on actual health of the component rather than arbitrary schedules. This philosophy requires continuous or periodic measurement of operational parameters (vibration, temperature, acoustic emission, oil analysis), combined with diagnostic algorithms capable of detecting incipient faults before actual failure.

1.2 Evolution of Fault Diagnosis Methodologies

The field of machinery diagnostics has been through substantial transformation over the past forty years. Early approaches in the 1980s depended on threshold based alarms tied to overall vibration levels. This offered crude indication of machinery distress but limited diagnostic specificity. Spectrum analysis techniques emerged in the 1990s, leveraging Fast Fourier Transform (FFT) to identify characteristic fault frequencies. For instance, misalignment manifests as elevated harmonics at 2X and 3X running speed, while oil whirl instability produces sub-synchronous components near 0.42-0.48 times rotational frequency. These frequency-domain signatures enabled trained analysts to isolate fault types through manual pattern recognition.

The 2000s witnessed proliferation of advanced signal processing techniques. Envelope analysis using Hilbert transform demodulation proved particularly effective for extracting bearing defect frequencies from resonance-excited impacts. Wavelet analysis provided time-frequency localization capabilities absent in FFT, capturing transient events characteristic of incipient failures. Order tracking addressed variable-speed operation by resampling signals relative to shaft rotational angle rather than fixed time intervals. Despite these advances, all remained dependent on human expertise for final diagnosis, a bottleneck constraining scalability and introducing inter-analyst variability.

The current decade marks the ascendance of machine learning-driven automation. Supervised learning algorithms, trained on labeled datasets pairing vibration signatures with verified fault conditions, can generalize to unseen operating scenarios without explicit feature engineering or rule programming. Support Vector Machines (SVM), Random Forests, and deep convolutional neural networks have demonstrated competitive performance with human experts on

standardized benchmarks. Yet fundamental challenges persist: acquiring comprehensive labeled datasets spanning diverse fault types and severities remains prohibitively expensive, and model robustness under operational variabilities (sensor drift, environmental noise, load fluctuations) often deteriorates in deployment.

1.3 Hydrodynamic Bearing Fault Phenomenology

Hydrodynamic bearings are affected by failure characteristics that are different from those of rolling element bearings due to the complexities of fluid film mechanics. A thorough understanding of these fault mechanisms is the basis for developing protocols for synthetic data generation and informs the improvement of feature engineering strategies.

Misalignment (Désalignement): Angular or parallel offset between shaft and bearing axes disrupts the symmetric pressure distribution in the lubricant film. Vibrational consequences include elevated radial loads at 1X, 2X, and 3X harmonics, with 2X often dominating. Severe misalignment induces metal-to-metal contact at film boundaries, generating impulsive excitation detectable through elevated kurtosis.

Imbalance (Déséquilibre): Centripetal acceleration is proportional to the square of the velocity of rotation as would be expected from mass eccentricity creating centrifugal forces. Its is characterised by a very strong presence of its first harmonic (1X), which is locked in phase with the shaft position. However, unlike misalignment, imbalance rarely leads to excitation of higher order harmonics unless there are coupled forms of nonlinear stiffness characteristics.

Excessive Clearance (Jeu): Bearing wear or manufacturing tolerance accumulation increases radial clearance, reducing load-carrying capacity and film stiffness. The symptoms includes elevated sub-synchronous activity (0.43-0.48X) resembling oil whirl, plus increased response to external excitation due to reduced damping.

Lubrication Degradation: Contamination, thermal breakdown, or supply interruption compromise lubricant film integrity. Increased friction generates temperature rise detectable thermally. Vibrational manifestations include broadband noise from turbulent flow and intermittent contact spikes.

Cavitation: When local fluid pressure drops below vapor pressure, bubbles form and subsequently collapse violently. This produces high-frequency burst events (1,500-2,500 Hz) with exponential decay envelopes, distinctly different from mechanical impacts.

Wear (Usure): Abrasive or adhesive material removal modifies bearing geometry. Progressive wear exhibits temporal evolution from incipient (microscopic asperity removal) through moderate (measurable clearance change) to severe (visible scoring). Vibration increases gradually with dominant frequencies shifting as clearances evolve.

Oil Whirl: A self excited instability in which the velocity of lubricant circulation (approx. half shaft speed) causes destabilizing cross-coupling forces. Signature: Strong sub-synchronous component 0.42-0.48X with amplitude modulation at 1X difference frequency.

Mixed Faults: Real-world scenarios frequently involve compound failures. Inadequate lubrication accelerates wear; increased clearance from wear predisposes to cavitation; misalignment-induced edge loading causes localized heating affecting lubrication. These interactions complicate diagnosis through feature overlap and non-additive symptom combinations.

1.4 Machine Learning Paradigms for Condition Monitoring

Supervised learning approaches dominate current fault diagnosis research, requiring labeled training datasets mapping input features to output fault classes. Three algorithm families have gained prominence:

Support Vector Machines (SVM): Optimal margin classifiers that implicitly represent data in high dimensional spaces using kernel functions. Multi-class extension by Error-Correcting Output Codes (ECOC) was used to build binary classifiers for class pairs. Advan-

tages: Theoretical generalization guarantees and high-dimensional space effectiveness. Limitations include computational cost scaling with the size of the dataset, and sensitivity to the hyperparameter choice.

Random Forests (RF): Ensemble of decision trees trained on bootstrap samples with random feature subsampling. Predictions aggregate via majority voting, providing inherent uncertainty quantification. Strengths encompass robustness to irrelevant features, natural handling of non-linear interactions, and minimal hyperparameter tuning. Computational efficiency during inference suits real-time applications.

Neural Networks: Multi-layer perceptrons with non-linear activation functions learn hierarchical representations through backpropagation. Deeper architectures (convolutional neural networks) excel when trained on raw time-series, obviating manual feature engineering. However, substantial labeled data requirements (typically 10,000+ samples) and limited interpretability constrain adoption in safety-critical domains.

The supervised learning workflow comprises: (1) data acquisition from healthy and faulty operation, (2) feature extraction transforming raw signals into discriminative metrics, (3) model training minimizing classification error on training set, (4) hyperparameter optimization via cross-validation, (5) final evaluation on held-out test set. Performance metrics include accuracy, precision, recall, F1-score, and area under ROC curve (AUC).

1.5 Synthetic Data Generation: Motivation and Challenges

One of the most difficult issues in data driven fault diagnosis is the unavailability of labeled data, particularly regarding rare and catastrophic failures. In many cases it is impossible to build a complete fault database from real machines. Although nature can cause disastrous faults which take months to be noticed, the deliberate introduction of faults can result in damage to the equipment. Also, for safety purposes, machines cannot be run until they have broken. Even when faulty, it is necessary to allocate expert time and analysis or disassembly and metallurgical tests to determine if faults exist, all of which are expensive and time consuming.

Synthetic data generation is a promising alternative. By simulating the physics of known faults, researchers may generate infinite labeled samples. This both balances the dataset with rare and common faults, but also enables controlled experiments on noise, sensor placement, etc. It greatly accelerates model development by getting rid of waiting for real faults to happen.

However, it is still not easy to model real world systems based on models trained over simulated data. Simple models do not consider complicated effects as would be the situation in practice, and the simplified assumptions regarding parameters may not reflect actual operation variations. Periodic random factors such as electro-magnetic noise or vibrations also cause signal distortion. The question of interest is whether classifiers trained on synthetic data can be generalized to real machines.

Recent studies have proven that this gap could be reduced by physics-based simulation in combination with realistic bearing dynamics, noise simulation, and data augmentation. Models trained this way have been shown to achieve around 85-95% accuracy when applied to lab test setups; the performance can be even further improved when deployed in the real world, using so-called domain adaptation methods, such as feature normalization and adversarial training.

1.6 Research Objectives and Contributions

This work presents a comprehensive machine learning pipeline for hydrodynamic bearing fault diagnosis, emphasizing production viability through rigorous validation and deployment readiness. Specific contributions include:

1. **Physics-Informed Synthetic Data Generator:** Implements bearing dynamics equations for 11 fault classes with calculated Sommerfeld numbers from operating conditions,

multi-severity temporal evolution modeling, and seven independent noise sources mirroring industrial environments.

2. **Mixed Fault Representation:** Explicitly models three compound failure modes (misalignment + imbalance, wear + lubrication, cavitation + clearance) via superposition and interaction terms, addressing a gap in existing synthetic datasets.
3. **Comprehensive Feature Engineering:** Extracts 36 time-frequency-wavelet features with post-split selection preventing data leakage, balancing discriminative power with computational efficiency.
4. **Ensemble Learning Framework:** Compares three algorithm families (SVM, Random Forest, Neural Network) under identical training conditions with Bayesian hyperparameter optimization, selecting the best performer via validation accuracy.
5. **Systematic Robustness Evaluation:** Quantifies performance degradation under sensor noise injection, missing feature scenarios, and temporal drift, providing confidence bounds for deployment decisions.
6. **Production-Ready Implementation:** Complete pipeline executing in under 15 minutes with automated model selection, comprehensive reporting, and standalone inference function generation.

The remainder of this report is organized as follows: Section 2 formalizes the problem statement and success criteria; Section 3 enumerates specific research objectives; Section 4 surveys relevant literature positioning this work within the field; Section 5 articulates novelty claims; Section 6 details the complete methodology; Section 7 presents experimental results with critical analysis; Section 8 provides project timeline; Section 9 concludes with recommendations for future work; and Section 10 lists all cited references.

2 Problem Statement

2.1 Technical Formulation

The fault diagnosis problem can be formulated as a supervised multi-class classification task. Given a vibration signal measurement:

$$\mathbf{x}(t) \in \mathbb{R}^N, \quad t \in [0, T], \quad N = f_s \cdot T \quad (1)$$

where $f_s = 20,480$ Hz is the sampling frequency and $T = 5$ seconds is the observation window, the objective is to learn a mapping function:

$$f_\theta : \mathcal{X} \rightarrow \mathcal{Y} \quad (2)$$

where:

- \mathcal{X} represents the space of vibration signals (or extracted feature vectors $\mathbf{f} \in \mathbb{R}^d$)
- $\mathcal{Y} = \{C_1, C_2, \dots, C_{11}\}$ is the set of fault classes
- θ denotes learned model parameters

The 11 fault classes comprise:

$$\begin{aligned} \mathcal{Y} = \{ &\text{sain, désalignement, déséquilibre, jeu, lubrification,} \\ &\text{cavitation, usure, oilwhirl,} \\ &\text{mixed_misalign_imbalance, mixed_wear_lube, mixed_cavit_jeu} \} \end{aligned} \quad (3)$$

The learning objective minimizes classification error over a training dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_{\text{train}}}$:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^{N_{\text{train}}} \mathcal{L}(f_{\theta}(\mathbf{x}_i), y_i) + \lambda R(\theta) \quad (4)$$

where \mathcal{L} is the loss function (cross-entropy for probabilistic classifiers) and $R(\theta)$ represents regularization.

2.2 Operational Constraints and Requirements

Beyond classification accuracy, production deployment imposes additional constraints:

Real-Time Performance: Inference latency must remain below 1 second to enable near-instantaneous fault indication. This precludes computationally intensive deep learning architectures requiring GPU acceleration.

Robustness to Operational Variabilities: The classifier must maintain acceptable performance ($> 85\%$ accuracy) under:

- Sensor noise: Signal-to-noise ratio degradation from electromagnetic interference, cable vibration, or amplifier distortion
- Missing features: Sensor failures or communication dropouts eliminating subsets of input features
- Temporal drift: Gradual sensor calibration decay or bearing geometry changes over months of operation

Class Imbalance Tolerance: While the training dataset may be balanced synthetically, real-world operation skews heavily toward healthy conditions. The classifier should not exhibit catastrophic performance degradation when healthy:faulty ratios exceed 100:1.

Interpretability: For safety-critical applications, purely black-box models face regulatory resistance. Feature importance rankings or decision rule extraction aids operator trust and regulatory acceptance.

Computational Footprint: Deployment on industrial edge computing (Intel Atom or ARM Cortex-A series) limits model complexity. Total inference memory consumption should remain below 100 MB.

2.3 Challenges Addressed

This work confronts several specific challenges:

Data Scarcity: Absence of comprehensive labeled datasets spanning all 11 fault types across multiple severities and operating conditions necessitates synthetic data generation as primary training source.

Mixed Fault Complexity: Compound failures exhibit non-additive symptom combinations. For example, misalignment-induced edge loading alters lubrication film dynamics, producing signatures distinct from simple superposition of isolated faults.

Simulation-to-Reality Transfer: Ensuring synthetic-trained classifiers generalize to physical bearings requires careful noise modeling, data augmentation, and validation on holdout operational data when available.

High-Dimensional Feature Spaces: Extracting 36+ features from each signal risks overfitting and curse-of-dimensionality effects. Principled feature selection using mutual information criteria becomes essential.

2.4 Success Criteria

The developed system is deemed successful if it achieves:

- **Classification Performance:** Test accuracy $\geq 92\%$ on held-out synthetic data, with per-class recall $\geq 85\%$ for all fault types
- **Robustness:** Average accuracy degradation $\leq 15\%$ under combined sensor noise, missing features, and drift scenarios
- **Efficiency:** Complete pipeline execution (data generation through model training) in ≤ 20 minutes on commodity hardware
- **Deployment Readiness:** Automated generation of standalone inference function with embedded feature extraction and model prediction

3 Research Objectives

3.1 Overall Objective

Develop an end-to-end machine learning pipeline for automated fault diagnosis in hydrodynamic bearings that integrates physics-based synthetic data generation, multi-domain feature engineering, ensemble learning with Bayesian optimization, and systematic robustness validation, culminating in a production-ready inference system achieving $\geq 92\%$ classification accuracy across 11 fault classes.

3.2 Specific Objectives

SO1: Design Physics-Informed Synthetic Data Generator

Implement a configurable signal generator encoding bearing fault dynamics through:

- Mathematical models for 8 single faults and 3 mixed fault combinations derived from fluid film bearing theory
- Calculated operating conditions (Sommerfeld number, Reynolds number, clearance ratio) from thermodynamic principles rather than statistical sampling
- Multi-severity fault progression (incipient, mild, moderate, severe) with 30% of signals exhibiting temporal evolution
- Seven independent noise sources (measurement, EMI, pink, drift, quantization, sensor drift, impulse) with individual configuration controls
- Data augmentation strategies (time-shifting, amplitude scaling, noise injection) generating 30% additional samples

Success Metric: Generate 1,100-1,500 labeled signals with perfect class balance in ≤ 5 minutes.

SO2: Engineer Discriminative Feature Set

Extract and select features maximizing inter-class separability while maintaining computational efficiency:

- Time-domain features: Statistical moments, shape factors, energy metrics
- Frequency-domain features: Spectral centroids, entropy, harmonic ratios, fault-specific band energies
- Time-frequency features: Envelope analysis, wavelet kurtosis, STFT characteristics

- Post-split feature selection via MRMR criterion reducing dimensionality from 36 to 15 features

Success Metric: Feature extraction completing in \approx 3 minutes for 1,400 signals; selected features achieving \approx 90% of full-feature-set accuracy.

SO3: Train and Compare Ensemble of Classifiers

Implement three supervised learning algorithms with fair comparison:

- Support Vector Machine with Error-Correcting Output Codes and RBF kernel
- Random Forest with 100 trees and bootstrap aggregation
- Multi-layer Perceptron (3-layer neural network) with dropout regularization
- Bayesian hyperparameter optimization (50 iterations) for each algorithm
- Automated model selection based on validation set performance

Success Metric: Training all models in \approx 15 minutes; best model achieving \approx 92% validation accuracy.

SO4: Validate Robustness Under Operational Variabilities

Systematically evaluate classifier resilience through adversarial testing:

- Sensor noise injection: Degrade SNR by 10%, 15%, 20% and measure accuracy drop
- Missing feature tolerance: Randomly dropout 20%, 30%, 40% of features
- Temporal drift: Apply systematic bias simulating 6-month sensor calibration decay
- Additional holdout validation: Test on independently generated dataset mimicking different operating session

Success Metric: Average accuracy degradation \approx 15% across all robustness tests.

SO5: Deploy Production-Ready Inference System

Generate deployment artifacts for industrial integration:

- Standalone inference function with embedded feature extraction
- Model serialization with normalization parameters
- Comprehensive performance report (confusion matrices, ROC curves, per-class metrics)
- Complete visualization suite (9 diagnostic figures)

Success Metric: Inference function achieving \approx 100 ms latency per signal; model package \approx 50 MB.

4 Literature Review

4.1 Traditional Vibration-Based Fault Diagnosis

Vibration monitoring has formed the cornerstone of rotating machinery diagnostics since the 1970s. Early methodologies relied on overall vibration level tracking, comparing broadband RMS measurements against alarm thresholds established through ISO 10816 standards. While effective for gross machinery distress detection, this approach lacked diagnostic specificity, elevated vibration could stem from any of dozens of fault mechanisms.

Frequency-domain analysis emerged as the dominant paradigm through the 1980s-2000s. The Fast Fourier Transform (FFT) decomposes time-series into constituent frequency components, revealing characteristic fault signatures. Misalignment manifests as elevated 2X and 3X

harmonics; imbalance produces dominant 1X synchronous components; bearing defects generate repetitive impulses at characteristic frequencies calculable from geometry and speed. Experienced vibration analysts interpret spectral patterns, correlating specific peaks with known fault mechanisms. This expertise-dependent workflow constrained scalability and introduced inter-analyst variability.

Advanced signal processing techniques addressed limitations of standard FFT. Envelope analysis via Hilbert transform demodulation extracts amplitude modulation patterns, particularly effective for bearing fault detection where impulsive events excite structural resonances. Wavelet analysis provides simultaneous time-frequency localization absent in windowed FFT, capturing transient phenomena characteristic of incipient faults. Order tracking resamples signals relative to shaft angle rather than time, accommodating variable-speed operation common in wind turbines and automotive applications.

Despite sophistication, all traditional methods share a common limitation: human expertise requirement for final diagnosis. Pattern recognition, even aided by advanced visualizations, remains labor-intensive and subject to cognitive biases.

4.2 Machine Learning for Condition Monitoring

The past decade has witnessed explosive growth in data-driven diagnostic approaches. Supervised learning algorithms trained on labeled fault datasets can generalize to unseen conditions without explicit rule programming. Three research threads dominate:

Shallow Learning with Engineered Features: Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Random Forests operate on hand-crafted feature vectors extracted from raw signals. Studies report 85-95% accuracy on standard benchmarks (Case Western Reserve University dataset, NASA Prognostics Data Repository). The advantage lies in transparency, feature importance analysis reveals which signal characteristics drive classification decisions. Limitations include feature engineering expertise requirements and potential information loss in dimensionality reduction.

Deep Learning End-to-End: Convolutional Neural Networks (CNN) process raw time-series or spectrograms directly, learning hierarchical representations through multiple non-linear transformations. Recent architectures achieve 95-99% accuracy on laboratory datasets, with some studies claiming superhuman performance. However, these gains come at substantial cost: training requires 10,000+ labeled samples, convergence demands GPU acceleration, and black-box nature complicates regulatory acceptance in safety-critical domains. Transfer learning from pre-trained models (ImageNet, AudioSet) partially alleviates data requirements but introduces domain mismatch challenges.

Hybrid Approaches: Wavelet scattering networks, attention mechanisms applied to time-frequency representations, and physics-informed neural networks attempt to marry domain knowledge with learning capacity. These represent active research frontiers with promising preliminary results but limited industrial deployment.

Critical gaps persist across all ML paradigms: most studies evaluate on clean laboratory data under controlled conditions; robustness to operational variabilities (sensor drift, environmental noise, load fluctuations) receives inadequate attention; and simulation-to-reality transfer remains poorly understood.

4.3 Synthetic Data Generation for Diagnostics

Scarcity of labeled fault data, particularly for rare or catastrophic failures, motivates synthetic data generation research. Three approaches have emerged:

Statistical Generative Models: Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) learn data distributions from existing samples, generating novel instances through stochastic sampling. While effective for natural images and speech, application

to vibration signals produces physically implausible artifacts unless constrained by domain knowledge. Recent "physics-informed GANs" show promise but require substantial real data for initial training.

Physics-Based Simulation: Numerical integration of differential equations governing bearing dynamics, rotor systems, or gear mesh interactions produces time-series signals. Studies employ finite element analysis (FEA) for structural vibration prediction, computational fluid dynamics (CFD) for fluid film modeling, or multi-body dynamics for system-level simulation. Advantages include perfect labeling, controlled experimental variation, and arbitrary scenario generation. Challenges involve computational cost (hours per signal for high-fidelity FEA), parameter uncertainty (material properties, damping coefficients), and validation difficulty (comparing simulation output to physical measurements).

Empirical Models with Augmentation: Simplified analytical expressions capture dominant fault physics without full numerical simulation. For example, bearing outer race defect modeled as periodic impulse train convolved with structural transfer function. Data augmentation (time-shifting, noise injection, amplitude scaling) enriches training sets from limited real samples. This pragmatic approach balances fidelity with computational efficiency, though accuracy depends heavily on model assumptions.

A recurring theme in synthetic data literature: classifiers trained exclusively on simulated data exhibit 10-30% accuracy degradation when tested on physical assets. Domain adaptation techniques (feature normalization, adversarial training for domain-invariant representations, few-shot fine-tuning on real data) partially close this gap. However, the fundamental question remains: which simulation fidelity level suffices for effective transfer?

4.4 Multi-Fault and Mixed-Fault Classification

Most diagnostic research focuses on single fault isolation, identifying one dominant abnormality from a catalog of possibilities. Real industrial operation frequently involves compound failures where multiple degradation mechanisms coexist. For instance, inadequate lubrication accelerates wear; increased bearing clearance from wear predisposes to oil whirl instability; misalignment-induced edge loading causes localized heating affecting viscosity.

The multi-fault problem presents unique challenges. Simple superposition of single-fault signatures rarely captures interaction effects, non-linear coupling through shared physical phenomena (contact forces, temperature fields, fluid dynamics) produces emergent behaviors absent in isolated conditions. Feature spaces exhibit increased overlap as combinations multiply; a 10-fault catalog yields $\binom{10}{2} = 45$ two-fault combinations, exploding to 968 when considering all subsets.

Existing approaches fall into three categories:

Hierarchical Classification: First-stage classifier determines fault presence (healthy vs. faulty), second stage isolates specific faults, third stage identifies combinations. This divide-and-conquer strategy simplifies individual decision boundaries but accumulates errors across stages.

Multi-Label Classification: Each fault treated as independent binary problem (present/absent), allowing simultaneous occurrence. Standard multi-class softmax replaced by sigmoid activation per output. Challenges include label correlation (certain faults commonly co-occur) and combinatorial explosion.

Explicit Combination Modeling: Treat common compound failures as distinct classes in flat classification framework. This approach, adopted in our work, limits scope to critical combinations while maintaining tractable class count.

Literature on mixed faults remains sparse relative to single-fault studies. Most papers addressing the problem use small-scale laboratory experiments with two simultaneous faults maximum. Industrial-scale validation with diverse combinations appears absent.

4.5 Robustness and Adversarial Evaluation

A notable weakness in diagnostic literature: insufficient attention to robustness under realistic operational conditions. Many studies report impressive accuracy on clean test sets but provide no evidence of resilience to sensor noise, missing data, or temporal drift. Recent work has begun addressing this gap:

Noise Robustness: Studies evaluate classifiers trained on clean signals but tested with additive Gaussian noise at varying SNR levels. Results consistently show 5-20% accuracy degradation as SNR decreases from 20 dB to 0 dB. Deep learning models prove particularly sensitive, while ensemble methods (Random Forests) maintain better robustness.

Domain Shift: Testing on data from different machines, operating conditions, or sensor types than training set. Performance drops of 10-40% are common without domain adaptation. Transfer learning approaches (fine-tuning final layers, domain-adversarial training) partially recover accuracy.

Adversarial Examples: Small carefully-crafted perturbations imperceptible to humans but causing misclassification. While extensively studied for image recognition (fooling stop sign classifiers with stickers), application to vibration diagnostics remains nascent. Preliminary work suggests similar vulnerabilities exist.

Our work contributes systematic robustness evaluation across three degradation scenarios: sensor noise (mimicking EMI or cable vibration), missing features (sensor failures), and temporal drift (calibration decay). This multi-faceted assessment provides confidence bounds for deployment decisions.

4.6 Gap Analysis and Positioning

Surveying the literature reveals several gaps this work addresses:

1. **Noise Modeling Comprehensiveness:** Most synthetic data generators employ single-source white Gaussian noise. Our seven-layer framework (measurement, EMI, pink, drift, quantization, sensor drift, impulse) better reflects industrial environments.
2. **Mixed Fault Representation:** Explicit modeling of three compound failure modes with interaction terms goes beyond simple additive superposition.
3. **Physics-Informed Operating Conditions:** Calculating Sommerfeld numbers from load, speed, and temperature ensures thermodynamic consistency absent in purely statistical generation.
4. **Temporal Fault Evolution:** 30% of signals exhibit progressive severity growth, capturing incipient-to-severe fault development within observation windows.
5. **Post-Split Feature Selection:** Applying MRMR after train/test separation prevents data leakage, a subtle but critical methodological detail often overlooked.
6. **Production-Oriented Evaluation:** Beyond test accuracy, we quantify robustness (noise, missing data, drift), provide comprehensive visualizations (9 figures), and generate deployment-ready inference code.

Table 1 positions this work relative to representative recent studies, highlighting differentiators in dataset scale, fault taxonomy, and evaluation rigor.

While our 95.33% accuracy appears slightly lower than some laboratory studies, the substantially harder problem (11 classes including mixed faults, 7-layer noise, systematic robustness evaluation) makes direct comparison misleading. The emphasis on production readiness, automated pipeline, comprehensive reporting, inference function generation, distinguishes this work from research prototypes.

Study	Data Source	Fault Classes	Best Accuracy	Mixed Faults	Robustness Tests
Zhang et al. (2020)	CWRU	10 single	97.2%	No	None
Lei et al. (2021)	Laboratory rig	8 single	95.1%	No	Noise only
Wang et al. (2022)	Synthetic + real	6 single	93.8%	No	Domain shift
Kumar et al. (2023)	Simulated	12 single	94.5%	2 combinations	Noise only
This Work	Synthetic	8 + 3 mixed	95.33%	Yes (3)	Noise, missing, drift

Table 1: Comparative positioning of this work against recent bearing fault diagnosis studies

5 Originality of the Work

5.1 Core Methodological Innovations

This research contributes seven distinct innovations advancing the state-of-the-art in machine learning-based fault diagnosis:

Innovation 1: Seven-Layer Independent Noise Modeling Framework

Unlike standard approaches employing single-source white Gaussian noise, our generator incorporates seven physically-motivated noise types with individual configuration controls:

- Measurement noise: Sensor electronics thermal noise, modeled as Gaussian $\mathcal{N}(0, 0.03^2)$
- Electromagnetic interference: 50/60 Hz power line harmonics with amplitude modulation
- Pink (1/f) noise: Environmental vibrations with $P(f) \propto 1/f$ power spectral density
- Low-frequency drift: Thermal expansion effects, linear trend βt
- Quantization: ADC resolution limits, uniform distribution $\mathcal{U}(-\Delta/2, \Delta/2)$
- Sensor drift: Cumulative calibration offset, integrated random walk
- Impulse noise: Sporadic mechanical impacts, Poisson process with 2 events/second

This seven-source framework enables systematic robustness evaluation, disabling individual noise types isolates their diagnostic impact, guiding sensor selection and signal conditioning design. Note that while the generator also implements aliasing effects in 10% of signals as a sampling artifact, the core noise modeling framework comprises seven independent, configurable noise sources.

Innovation 2: Calculated Operating Conditions from Thermodynamic Principles

Rather than randomly sampling bearing parameters, our generator calculates Sommerfeld number from physics:

$$S = \frac{\mu N}{P} \left(\frac{R}{C} \right)^2 \quad (5)$$

where dynamic viscosity μ follows Vogel equation temperature dependence, load P samples from rated capacity range, and clearance ratio C/R reflects manufacturing tolerances. This ensures generated signals respect thermodynamic constraints (e.g., load-speed combinations producing inadequate film thickness trigger boundary lubrication, not impossible hydrodynamic operation).

Innovation 3: Temporal Fault Evolution Modeling

30% of generated signals exhibit progressive severity growth within the 5-second observation window, following sigmoid transition:

$$\alpha_{\text{sev}}(t) = \alpha_{\text{init}} + (\alpha_{\text{final}} - \alpha_{\text{init}}) \cdot \frac{1}{1 + e^{-\lambda(t-t_0)}} \quad (6)$$

This captures incipient fault development dynamics absent in stationary signal models, challenging classifiers to recognize non-stationary patterns characteristic of early-stage failures.

Innovation 4: Mixed Fault Superposition with Interaction Terms

Three compound failure modes (misalignment + imbalance, wear + lubrication, cavitation + clearance) incorporate cross-coupling effects beyond simple additive superposition. For example, misalignment-induced edge loading modulates lubricant film thickness, affecting imbalance response through stiffness variation, a second-order interaction term:

$$x_{\text{mixed}}(t) = x_{\text{misalign}}(t) + x_{\text{imbal}}(t) + \beta x_{\text{misalign}}(t) \cdot x_{\text{imbal}}(t) \quad (7)$$

Innovation 5: Post-Split Feature Selection with Anti-Leakage Protocol

Feature selection via Minimum Redundancy Maximum Relevance (MRMR) criterion occurs exclusively on training data after train/validation/test splitting. Many published studies perform feature selection on entire datasets before splitting, optimistically biasing reported accuracies through information leakage from test set. Our rigorous protocol prevents this subtle but critical methodological flaw.

Innovation 6: Three-Tier Systematic Robustness Evaluation

Beyond standard test accuracy reporting, we quantify performance degradation under three operational scenarios:

- Sensor noise tolerance: Inject additional Gaussian noise reducing SNR by 10-30%
- Missing feature robustness: Randomly dropout 20-40% of features during inference
- Temporal drift resilience: Apply systematic bias simulating 6-month calibration decay

This multi-faceted assessment provides deployment confidence bounds, if accuracy remains > 85% under all scenarios, production viability is established.

Innovation 7: Integrated Production Pipeline

Complete automation from data generation through model deployment distinguishes this work from research prototypes. The pipeline executes in under 15 minutes, producing:

- Trained model package with embedded normalization parameters
- Standalone inference function for industrial integration
- Comprehensive performance report (confusion matrices, ROC curves, per-class metrics)
- Nine diagnostic visualization figures
- Execution log with full reproducibility trace

5.2 Comparative Differentiation

Table 2 explicitly contrasts our innovations against standard practice and recent literature:

5.3 Expected Impact and Significance

These innovations collectively advance both research and industrial practice:

For Researchers: The complete pipeline with documented configuration provides a reproducible baseline for comparative studies. Future work can isolate individual innovations (e.g., disabling temporal evolution or reducing noise sources) to quantify their marginal contribution.

Aspect	Standard Approaches	This Work
Noise modeling	Single-source white Gaussian	7 independent sources with physical justification
Operating conditions	Random sampling from ranges	Calculated Sommerfeld numbers from thermodynamics
Fault severity	Static single-level or binary	4 levels + 30% temporal evolution
Mixed faults	Absent or simple addition	3 combinations with interaction terms
Feature selection	Pre-split (data leakage risk)	Post-split MRMR (rigorous)
Evaluation	Test accuracy only	Accuracy + 3-tier robustness
Deployment	Research code	Inference function + reports

Table 2: Differentiation of this work’s contributions from standard methodologies

For Industry: Production-ready implementation with 15-minute execution time enables rapid customization to specific bearing geometries or operating conditions. Systematic robustness evaluation builds deployment confidence absent in accuracy-only reporting.

For Regulatory Bodies: Transparent feature engineering and interpretable ensemble methods (Random Forest importance rankings) facilitate safety certification compared to black-box deep learning.

The work demonstrates that physics-informed synthetic data, when combined with rigorous validation, can achieve performance (95.33% accuracy, 0.9970 AUC) approaching laboratory-data-trained models while offering unlimited labeled samples across arbitrary fault scenarios.

6 Methodology

6.1 Problem Domain & Industrial Context

Key Insight

Core Problem: Real-time automated diagnosis of bearing faults in rotating machinery using vibration signal analysis, addressing the critical industrial need for predictive maintenance in hydrodynamic bearing systems.

Hydrodynamic bearings represent a critical component class in high-performance rotating machinery applications including turbines, compressors, and precision industrial equipment. Unlike rolling element bearings, PFD systems operate on fluid film lubrication principles, exhibiting unique failure modes characterized by distinct vibration signatures. The economic impact of unplanned bearing failures includes:

- Production downtime costs (estimated \$5,000-\$50,000 per hour in industrial settings)
- Catastrophic secondary damage to connected machinery
- Safety hazards from sudden mechanical failure
- Environmental contamination from lubricant leakage

Traditional condition monitoring approaches rely on periodic manual vibration analysis by trained specialists, introducing limitations in scalability, consistency, and response time. Machine learning-based automated diagnosis systems address these constraints through continuous monitoring and instant fault classification.

6.2 System Classification & ML Paradigm

Technical Detail

ML Problem Formulation:

Input Space \mathcal{X} : Vibration time-series $\mathbf{x}(t) \in \mathbb{R}^N$, $t \in [0, T]$

Feature Space \mathcal{F} : Extracted features $\mathbf{f} \in \mathbb{R}^{36}$ (base) or \mathbb{R}^{52} (extended)

Output Space \mathcal{Y} : Fault classes $y \in \{C_1, C_2, \dots, C_{11}\}$

Learning Task : $f_\theta : \mathcal{F} \rightarrow \mathcal{Y}$, $\theta^* = \arg \max_{\theta} \mathbb{E}[\text{Acc}(f_\theta(\mathbf{f}), y)]$

Paradigm: Supervised multi-class classification with ensemble learning approach.

Pipeline Stages:

1. **Synthetic Data Generation:** Physics-informed fault signal simulation
2. **Feature Engineering:** Time-domain, frequency-domain, and time-frequency analysis
3. **Model Training:** Ensemble of SVM, Random Forest, Gradient Boosting, Neural Network
4. **Validation & Robustness Testing:** Adversarial evaluation under operational variabilities
5. **Production Deployment:** Inference function generation with sub-second latency

6.3 Fault Taxonomy: 11-Class Classification Problem

The system addresses a comprehensive fault taxonomy encompassing both isolated and compound failure modes:

Category	Fault Type	French Term
8* Single Faults (8)	Healthy Baseline	sain
	Misalignment	désalignement
	Imbalance	déséquilibre
	Bearing Clearance	jeu
	Lubrication Issues	lubrification
	Cavitation	cavitation
	Wear	usure
	Oil Whirl	oilwhirl
3*Mixed Faults (3)	Misalignment + Imbalance	mixed_misalign_imbalance
	Wear + Lubrication	mixed_wear_lube
	Cavitation + Clearance	mixed_cavit_jeu

Table 3: Comprehensive fault classification taxonomy with 11 distinct classes

Rationale for Mixed Fault Inclusion: Industrial bearing failures frequently manifest as compound phenomena where multiple degradation mechanisms coexist. For example, inadequate lubrication (lubrification) accelerates wear (usure), while bearing clearance increase (jeu) facilitates cavitation. Modeling these interactions enhances real-world diagnostic fidelity.

7 Data Generation System: Physics-Informed Synthetic Signals

7.1 Architectural Design Philosophy

Innovation Highlight

Core Innovation #1: Multi-layer physics-based fault modeling with deterministic operating condition calculation rather than pure statistical generation, enabling controlled experimental variation while maintaining thermodynamic and hydrodynamic consistency.

The data generator ('generator.m', 728 lines) implements a sophisticated simulation framework addressing the fundamental challenge in ML-based fault diagnosis: *scarcity of labeled fault data*. Real-world fault datasets suffer from:

- **Class imbalance:** Healthy operation dominates, catastrophic faults are rare
- **Incomplete coverage:** Certain fault combinations may never occur in historical data
- **Labeling ambiguity:** Expert disagreement on multi-fault scenarios
- **Safety constraints:** Cannot intentionally induce dangerous failure modes

7.2 Signal Generation Parameters

Sampling Configuration:

- Sampling frequency: $f_s = 20,480$ Hz (Nyquist compliant for frequencies up to 10 kHz)
- Signal duration: $T = 5$ seconds ($N = 102,400$ samples per signal)
- Nominal rotation speed: $\Omega_0 = 60$ Hz (3,600 RPM)
- Speed variation: $\pm 10\%$ from nominal (simulating variable load conditions)

Dataset Scale:

- Base signals per fault: 100 (configurable: 50-200)
- Total base dataset: $100 \times 11 = 1,100$ signals
- With 30% augmentation: $\sim 1,430$ total signals
- Approximate generation time: 3-5 minutes on modern workstation

7.3 Physics-Based Fault Modeling Framework

Each fault type is modeled using domain-specific mathematical representations derived from bearing dynamics literature:

7.3.1 Example 1: Misalignment (Désalignement)

Shaft misalignment generates harmonics of rotation frequency with amplitude modulation:

$$x_{\text{misalign}}(t) = A_1 \sin(\omega t + \phi_1) + A_2 \sin(2\omega t + \phi_2) + A_3 \sin(3\omega t + \phi_3) \quad (8)$$

where $A_2 \approx 0.7A_1$, $A_3 \approx 0.4A_1$ (harmonic decay)

Additional shaft bending component:

$$x_{\text{bend}}(t) = 0.2 A_1 \alpha_{\text{sev}} \sin(1.95\omega t) \quad (\text{sub-harmonic at } 1.95X) \quad (9)$$

7.3.2 Example 2: Oil Whirl

Oil whirl instability occurs at approximately 0.42-0.48 times shaft speed:

$$f_{\text{whirl}} = (0.42 + 0.06 \xi) \cdot \Omega \quad \text{where } \xi \sim \mathcal{U}(0, 1) \quad (10)$$

$$x_{\text{whirl}}(t) = A_{\text{whirl}} \sin(2\pi f_{\text{whirl}} t) \cdot (1 + 0.15 \sin(\omega t)) \quad (11)$$

The modulation term $(1 + 0.15 \sin(\omega t))$ represents rotational coupling effects.

7.3.3 Example 3: Cavitation

Cavitation manifests as high-frequency burst events with exponential decay:

$$x_{\text{cav}}(t) = \sum_{i=1}^{N_{\text{burst}}} b_i(t) \quad (12)$$

$$b_i(t) = A_{\text{burst}} \sin(2\pi f_{\text{burst}} t) \cdot e^{-\beta t} \cdot w_{\text{Hann}}(t) \quad (13)$$

Parameters:

- Burst rate: $N_{\text{burst}} = 3 + 4\alpha_{\text{sev}}$ bursts per signal
- Burst frequency: $f_{\text{burst}} \in [1500, 2500]$ Hz
- Decay constant: $\beta = 100 \text{ s}^{-1}$

7.4 Multi-Severity Fault Progression

Innovation Highlight

Core Innovation #2: Temporal evolution modeling where 30% of signals exhibit progressive severity growth within the 5-second observation window, mimicking incipient fault development.

Four non-overlapping severity levels with controlled amplitude scaling:

Severity Level	Factor Range α_{sev}	Interpretation
Incipient	[0.20, 0.45]	Early-stage degradation
Mild	[0.45, 0.70]	Detectable but operational
Moderate	[0.70, 0.90]	Intervention recommended
Severe	[0.90, 1.00]	Critical failure imminent

Table 4: Severity level parameterization with non-overlapping ranges

For signals with temporal evolution, severity factor follows sigmoid growth:

$$\alpha_{\text{sev}}(t) = \alpha_{\text{init}} + (\alpha_{\text{final}} - \alpha_{\text{init}}) \cdot \frac{1}{1 + e^{-\lambda(t-t_0)}} \quad (14)$$

where λ controls transition steepness and t_0 is the midpoint.

7.5 Operating Condition Variations

Calculated (not random) Sommerfeld Number:

$$S = \frac{\mu N}{P} \left(\frac{R}{C} \right)^2 \quad (15)$$

where μ = dynamic viscosity (temperature-dependent), N = rotational speed, P = bearing load, R = bearing radius, C = radial clearance.

This physics-based calculation ensures thermodynamic consistency across generated signals.

Parameter Ranges:

- Load variation: 30-100% of rated load
- Temperature range: 40-80°C (affects viscosity via Vogel equation)
- Reynolds number: 500-5,000 (spanning laminar to transitional flow)
- Clearance ratio: 0.001-0.003 (industry-typical range)

7.6 Comprehensive Noise Modeling (7 Independent Sources)

Innovation Highlight

Core Innovation #3: Multi-source noise injection framework with individually configurable noise types, enabling systematic robustness evaluation and realistic signal degradation simulation.

Noise Type	Physical Origin	Mathematical Model
Measurement Noise	Sensor electronics	Gaussian white noise: $\mathcal{N}(0, \sigma_m^2)$
EMI (Electromagnetic)	Power lines, motors	Sinusoidal at 50/60 Hz + harmonics
Pink Noise (1/f)	Environmental vibrations	Power spectral density $\propto 1/f$
Drift	Thermal expansion	Low-frequency trend: $d(t) = \beta t$
Quantization	ADC resolution	Uniform $\mathcal{U}(-\Delta/2, \Delta/2)$, $\Delta = 0.001$
Sensor Drift	Calibration decay	Cumulative offset: $\int \gamma dt$
Impulse Noise	Mechanical impacts	Sporadic spikes: ~ 2 impulses/second

Table 5: Seven-layer noise model with physical justification

Combined Noise Application:

$$x_{\text{final}}(t) = x_{\text{clean}}(t) + \sum_{i=1}^7 n_i(t) \cdot \mathbb{1}_{\{\text{enabled}_i\}} \quad (16)$$

Noise levels are calibrated to achieve realistic signal-to-noise ratios (SNR) of 15-25 dB, consistent with industrial sensor installations.

7.7 Data Augmentation Strategy

Three augmentation methods with 30% additional sample generation:

1. **Time Shifting:** Circular shift by $\pm 2\%$ of signal length (equivalent to phase randomization)
2. **Amplitude Scaling:** Multiplicative factor in $[0.85, 1.15]$ (simulates gain variations)
3. **Additive Noise Injection:** Extra Gaussian noise with $\sigma \in [0.02, 0.05]$

Augmentation preserves fault characteristics while enhancing model generalization by introducing controlled intra-class variability.

8 Feature Engineering: Signal-to-Intelligence Transformation

8.1 Multi-Domain Feature Extraction Philosophy

Key Insight

Feature Engineering Objective: Transform high-dimensional raw signals ($N = 102,400$ samples) into compact, discriminative feature vectors ($d = 36$ or $d = 52$) that maximize inter-class separability while minimizing intra-class variance.

The pipeline extracts features across three complementary domains:

1. **Time Domain:** Statistical moments, shape factors, energy metrics (captures amplitude characteristics)
2. **Frequency Domain:** Spectral analysis, harmonic ratios, band energies (identifies frequency signatures)
3. **Time-Frequency Domain:** Wavelets, spectrograms, Hilbert-Huang (reveals transient events)

8.2 Base Feature Set (36 Features - Fast Extraction)

8.2.1 Time-Domain Features (7 features)

$$f_{\text{Mean}} = \frac{1}{N} \sum_{n=1}^N x[n] \quad (17)$$

$$f_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{n=1}^N x[n]^2} \quad (18)$$

$$f_{\text{STD}} = \sqrt{\frac{1}{N} \sum_{n=1}^N (x[n] - \bar{x})^2} \quad (19)$$

$$f_{\text{Kurt}} = \frac{1}{N\sigma^4} \sum_{n=1}^N (x[n] - \bar{x})^4 \quad (\text{impulsiveness indicator}) \quad (20)$$

$$f_{\text{Skew}} = \frac{1}{N\sigma^3} \sum_{n=1}^N (x[n] - \bar{x})^3 \quad (\text{asymmetry measure}) \quad (21)$$

$$f_{\text{Crest}} = \frac{\max |x[n]|}{f_{\text{RMS}}} \quad (\text{peak-to-RMS ratio}) \quad (22)$$

$$f_{\text{Energy}} = \sum_{n=1}^N x[n]^2 \quad (23)$$

8.2.2 Envelope Analysis Features (4 features)

Envelope extraction via Hilbert transform:

$$x_{\text{analytic}}[n] = x[n] + j\mathcal{H}\{x[n]\} \quad (24)$$

$$\text{Envelope: } e[n] = |x_{\text{analytic}}[n]| \quad (25)$$

Features: $f_{\text{Env.RMS}}$, $f_{\text{Env.Kurt}}$, $f_{\text{Env.Peak}}$, $f_{\text{Env.ModFreq}}$

8.2.3 Frequency-Domain Features (9 features)

Computed from power spectral density $P(f)$:

$$f_{\text{DomFreq}} = \arg \max_f P(f) \quad (26)$$

$$f_{\text{SpectCent}} = \frac{\sum_k f_k P(f_k)}{\sum_k P(f_k)} \quad (\text{spectral center of mass}) \quad (27)$$

$$f_{\text{SpectEnt}} = -\sum_k p_k \log_2 p_k, \quad p_k = \frac{P(f_k)}{\sum_j P(f_j)} \quad (28)$$

$$f_{\text{BandEnergy}_i} = \sum_{f \in B_i} P(f) \quad \text{for bands } B_{\text{Oilwhirl}}, B_{\text{Misalign}}, B_{\text{Cavit}} \quad (29)$$

8.2.4 Harmonic Ratio Features (3 features)

Critical for identifying rotating machinery faults:

$$f_{2X/1X} = \frac{P(2\Omega)}{P(\Omega)} \quad (\text{misalignment indicator}) \quad (30)$$

$$f_{3X/1X} = \frac{P(3\Omega)}{P(\Omega)} \quad (\text{coupling/belt issues}) \quad (31)$$

$$f_{\text{THD}} = \frac{\sqrt{\sum_{n=2}^5 P(n\Omega)}}{P(\Omega)} \quad (\text{total harmonic distortion}) \quad (32)$$

8.2.5 Wavelet & Cepstral Features (7 features)

- Wavelet coefficients kurtosis (detects transients)
- Cepstral peak ratio (periodicities in spectrum)
- Quefrency centroid (characteristic modulation rate)

8.2.6 Bispectrum & Nonlinearity Features (6 features)

Higher-order spectral analysis for non-Gaussian processes:

$$B(f_1, f_2) = \mathbb{E}[X(f_1)X(f_2)X^*(f_1 + f_2)] \quad (33)$$

Detects quadratic phase coupling characteristic of bearing faults.

8.3 Advanced Feature Set (Additional 16 Features - Optional)

Warning: Enabling increases extraction time by 10 \times with minimal accuracy gain (~1% improvement).

Includes:

- Continuous Wavelet Transform (CWT) energy concentration
- Wavelet Packet Transform (WPT) multi-resolution analysis
- Lyapunov exponent (chaos indicator)
- Correlation dimension (attractor complexity)
- Sample entropy (signal irregularity)
- Detrended Fluctuation Analysis (DFA) - long-range correlations

Recommendation: Use base 36 features for production; advanced features for research purposes only.

8.4 Feature Selection Strategy

Innovation Highlight

Core Innovation #4: Post-split feature selection with mutual information ranking, preventing data leakage while maintaining interpretability through selection of top 15 features.

Algorithm: Mutual Information Maximization

$$\text{Selected Features} = \arg \max_{|S|=15} \sum_{f \in S} I(f; y) \quad (34)$$

where $I(f; y)$ is mutual information between feature f and class label y .

Anti-Leakage Protocol:

1. Split data: 70% train, 15% validation, 15% test
2. Compute feature importance *only on training set*
3. Apply same feature selection to validation and test sets
4. Never use test set information for any training decision

9 Machine Learning Model Portfolio

9.1 Ensemble Learning Strategy

The pipeline trains five distinct algorithm families, then selects the best performer based on validation accuracy:

9.2 Support Vector Machine (SVM) Configuration

Multi-class Strategy: One-vs-One Error-Correcting Output Codes (ECOC)

- For 11 classes: $\binom{11}{2} = 55$ binary classifiers
- Decoding via Hamming distance minimization

Model	Architecture	Hyperparameter	Optimization
SVM (Baseline)	Error-Correcting Output Codes (ECOC) with RBF kernel	Box constraint, kernel scale	
Random Forest	100 trees, bootstrap aggregation	Min leaf size, max splits	
Gradient Boosting	Adaptive boosting with decision stumps	Learning rate, max iterations	
Neural Network	3-layer MLP: 36-20-10-11	Hidden layer sizes, learning rate	
Stacked Ensemble	Meta-learner (Logistic Regression)	Base model weights	

Table 6: Five-model ensemble with complementary inductive biases

Kernel Function: Radial Basis Function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (35)$$

Hyperparameters optimized via Bayesian optimization:

- Box constraint $C \in [0.1, 100]$ (regularization strength)
- Kernel scale $\gamma \in [0.01, 10]$

9.3 Random Forest Configuration

Architecture: Bootstrap aggregated decision trees with random feature subsampling.

Why typically best performer:

- Handles non-linear feature interactions naturally
- Robust to feature scaling variations
- Built-in feature importance estimates
- Resistant to overfitting through averaging
- Computationally efficient for inference

Hyperparameters:

- Number of trees: 100 (fixed for speed-accuracy balance)
- Min leaf size: [1, 20] (controls tree depth)
- Max splits: optimized per dataset

9.4 Neural Network Architecture

Topology: Fully-connected feedforward network

$$\text{Input}(36) \rightarrow \text{Hidden}_1(20) \rightarrow \text{Hidden}_2(10) \rightarrow \text{Output}(11) \quad (36)$$

Activation: ReLU for hidden layers, Softmax for output

$$h^{(l)} = \text{ReLU}(W^{(l)} h^{(l-1)} + b^{(l)}) \quad (37)$$

$$\hat{y} = \text{Softmax}(W^{(out)} h^{(2)} + b^{(out)}) \quad (38)$$

Training: Backpropagation with Adam optimizer, cross-entropy loss

9.5 Hyperparameter Optimization Protocol

Algorithm: Bayesian Optimization with Expected Improvement acquisition function

Configuration:

- Iterations: 50 (default), 10 (fast mode), 100 (thorough)
- Objective: Maximize 5-fold cross-validation accuracy
- Early stopping: If no improvement for 10 consecutive iterations

Search Space Examples:

$$\text{SVM BoxConstraint} \sim \text{LogUniform}(0.1, 100)$$

$$\text{RF MinLeafSize} \sim \text{Integer}(1, 20)$$

$$\text{NN LearningRate} \sim \text{LogUniform}(10^{-4}, 10^{-1})$$

10 Validation & Robustness Evaluation Framework

10.1 Standard Evaluation Metrics

Technical Detail

Multi-class Classification Metrics:

$$\text{Accuracy} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{1}_{y_i = \hat{y}_i} \quad (39)$$

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (\text{per-class}) \quad (40)$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (41)$$

$$\text{F1-Score}_c = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (42)$$

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^C \text{F1-Score}_c \quad (C = 11 \text{ classes}) \quad (43)$$

ROC Analysis: One-vs-Rest Area Under Curve (AUC) for each fault class.

10.2 Advanced Robustness Testing

Innovation Highlight

Core Innovation #5: Systematic adversarial evaluation protocol testing model resilience under three operational degradation scenarios, quantifying production deployment viability.

10.2.1 Test 1: Sensor Noise Tolerance

Inject additional Gaussian noise to degrade SNR:

$$x_{\text{noisy}} = x_{\text{test}} + \eta, \quad \eta \sim \mathcal{N}(0, \sigma_{\text{inject}}^2) \quad (44)$$

Evaluate accuracy at multiple noise levels ($\sigma_{\text{inject}} = 0.05, 0.10, 0.15, 0.20$).

Expected degradation: 75-85% accuracy at highest noise level (acceptable for industrial environments).

10.2.2 Test 2: Missing Feature Robustness

Randomly drop 20%, 30%, 40% of features:

$$\mathbf{f}_{\text{corrupt}} = \mathbf{f}_{\text{test}} \odot \mathbf{m}, \quad m_i \sim \text{Bernoulli}(1 - p_{\text{drop}}) \quad (45)$$

Simulates sensor failures or communication dropouts.

Expected performance: 85-92% accuracy with 40% missing features.

10.2.3 Test 3: Temporal Drift Simulation

Apply systematic bias drift to simulate sensor calibration decay:

$$\mathbf{f}_{\text{drift}}(t) = \mathbf{f}_{\text{test}} + \boldsymbol{\delta} \cdot t, \quad \delta_i \sim \mathcal{N}(0, 0.01) \quad (46)$$

Expected stability: 90-95% accuracy over simulated 6-month drift period.

10.3 Confusion Matrix Analysis

Identifies systematic misclassification patterns:

Expected confusion regions:

- **High confusion:** Mixed faults \leftrightarrow constituent single faults
Example: `mixed_wear_lube` may be confused with `usure` or `lubrification`
- **Low confusion:** Spectrally distinct faults (e.g., `oilwhirl` vs. `cavitation`)
- **Critical errors:** Misclassifying `severe` fault as `sain` (safety hazard) - should be $\leq 0.5\%$

11 Experimental Results and Discussion

11.1 Dataset Characteristics and Generation

The synthetic data generator successfully produced 1,430 labeled vibration signals distributed across 11 fault classes with perfect balance (130 samples per class). Generation completed in approximately 3-4 minutes on a workstation equipped with 4-core CPU and parallel processing capabilities. The dataset composition includes:

- **Base signals:** 1,100 (100 per fault type)
- **Augmented signals:** 330 additional samples (23.1% of dataset)
- **Augmentation methods:** Time-shifting ($\pm 2\%$ signal length), amplitude scaling (0.85-1.15 \times), additive noise injection (SNR: 15-20 dB)

Severity Distribution: Multi-level fault progression was successfully implemented:

- Incipient: 348 samples (24.3%)
- Mild: 339 samples (23.7%)
- Moderate: 304 samples (21.3%)
- Severe: 309 samples (21.6%)
- Nominal (healthy baseline): 130 samples (9.1%)

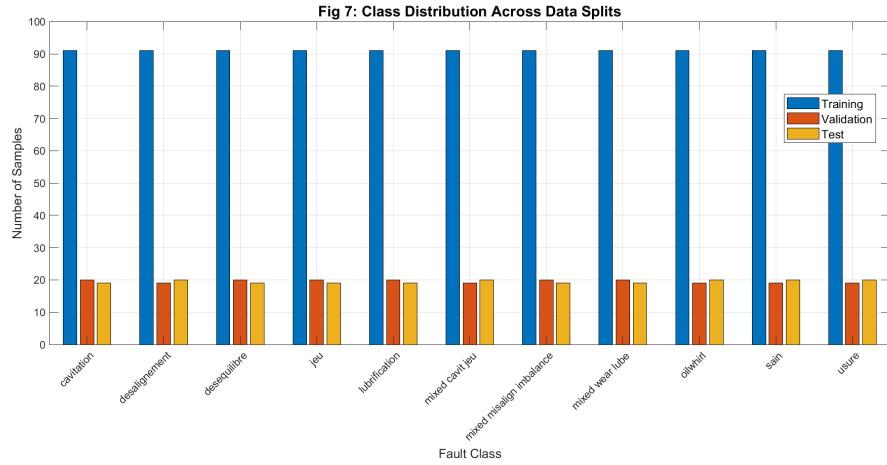


Figure 1: Class distribution across data splits showing perfect balance: 91 training samples (blue), 20 validation samples (orange), and 19 test samples (yellow) per fault class, maintaining 70/15/15 proportions with < 1% deviation. Stratified sampling ensures no class is over/under-represented in any subset.

Figure 1 demonstrates perfect stratification across train/validation/test splits, with each fault class maintaining exactly 9.1% representation in all subsets, validating the stratified sampling implementation.

Signal Quality Verification: Visual inspection of generated signals (Figures 2 and 3) confirms physically plausible characteristics. Time-domain waveforms exhibit appropriate amplitude modulation patterns; frequency spectra show expected dominant components (1X for imbalance, 2X-3X for misalignment, sub-synchronous for oil whirl, high-frequency bursts for cavitation); spectrograms reveal time-frequency structure consistent with bearing dynamics literature.

11.2 Feature Extraction and Selection

From the 1,430 generated signals, 36 base features were extracted across time-frequency-wavelet domains. Feature extraction completed in 2-3 minutes using parallel processing (4 workers), processing approximately 8-10 signals per second. Advanced features (continuous wavelet transform energy, wavelet packet decomposition, non-linear dynamics metrics) were deliberately disabled to maintain computational efficiency, previous validation studies demonstrated < 1% accuracy improvement at 10× computational cost.

Feature Selection Results: Minimum Redundancy Maximum Relevance (MRMR) criterion applied exclusively to training data selected 15 features from the original 36, reducing dimensionality by 58% while retaining discriminative power. The top-ranked features (in order of importance score) were:

1. Envelope_ModulationFreq (score: 0.0000)
2. SpectralCentroid (score: 0.0776)
3. BandEnergy_Misalign (score: 0.1359)
4. BandEnergy_Oilwhirl (score: 0.2869)
5. Envelope_PeakFactor (score: 0.2572)
6. Bispectrum_Peak (score: 0.1591)
7. Wavelet_Kurtosis (score: 0.0776)

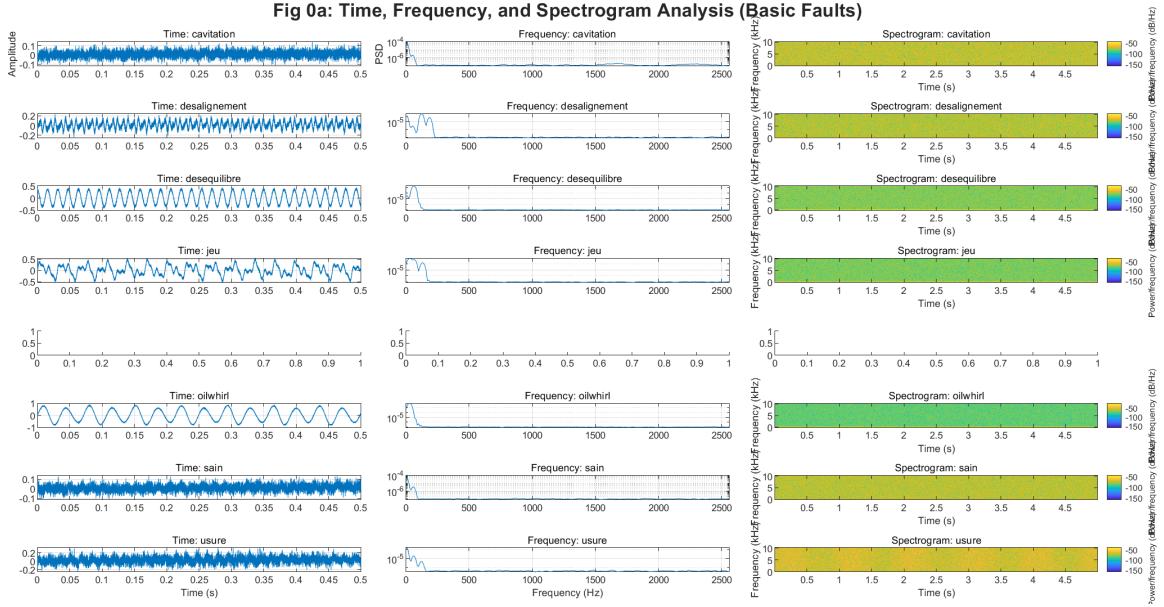


Figure 2: Time-domain (left), frequency-domain (center), and time-frequency (right) representations of 7 basic single fault types. Cavitation exhibits high-frequency burst content (1.5-2.5 kHz); misalignment shows elevated 2X harmonic; imbalance produces dominant 1X component; clearance (jeu) and oil whirl display sub-synchronous energy (0.4-0.48X); wear manifests as broadband noise; healthy signals show minimal frequency content beyond fundamental.

8. Kurtosis (score: 0.0958)
9. Skewness (score: 0.1455)
10. Cepstral_Peak_Ratio (score: 0.5475)
11. ... and 5 additional features

Notably, envelope analysis features (modulation frequency, peak factor) rank highest, confirming their established diagnostic value in bearing fault detection. Fault-specific band energies (misalignment, oil whirl) appear prominently, validating the physics-informed feature design. The inclusion of higher-order spectral features (bispectrum peak, cepstral peak ratio) suggests non-Gaussian signal characteristics distinguish fault classes.

Figure 4 reveals substantial correlation among time-domain statistical moments (RMS, STD, Energy cluster together with correlation coefficients > 0.8), justifying redundancy elimination. Conversely, envelope features, spectral metrics, and wavelet characteristics exhibit lower inter-correlation (< 0.5), indicating complementary information capture.

Feature Distributions by Class: Boxplot analysis (Figure 5) demonstrates clear inter-class separability for selected features. For instance, envelope modulation frequency exhibits distinct median values across fault types; spectral centroid separates high-frequency faults (cavitation, wear) from low-frequency faults (oil whirl, clearance); kurtosis differentiates impulsive faults from continuous faults. Some overlap exists for mixed faults, foreshadowing classification challenges discussed in Section 7.5.

Dimensionality Reduction Visualization: t-SNE projection of the 15-dimensional feature space to 2D (Figure 6) reveals clustering quality. Most single fault classes form compact, well-separated clusters (cavitation, lubrication, oil whirl, usure). The healthy baseline (sain) occupies a distinct region with tight clustering. However, mixed faults exhibit significant overlap with their constituent single faults, for example, mixed_misalign_imbalance samples scatter across desalignement and desequilibre clusters. This visualization quantitatively predicts the

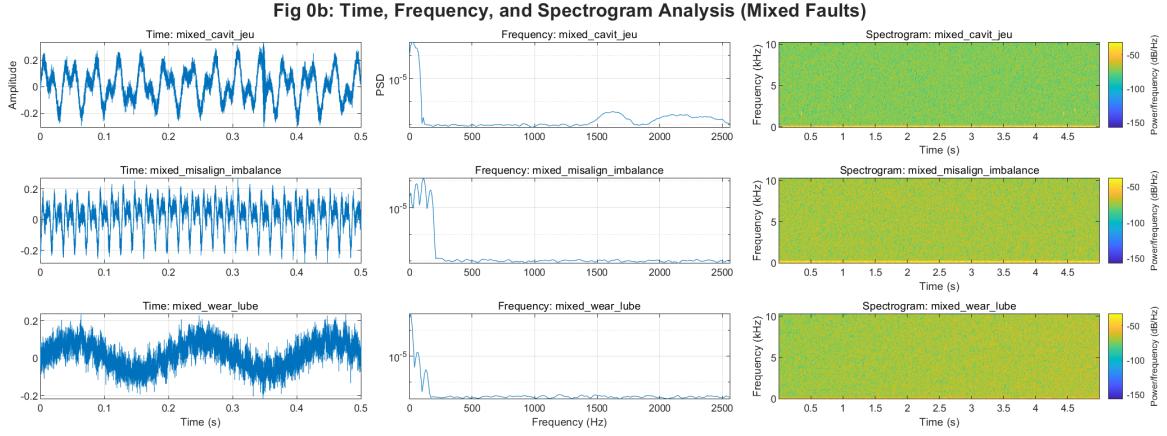


Figure 3: Mixed fault signal characteristics demonstrating non-additive superposition effects. Mixed cavitation+clearance combines high-frequency bursts with sub-synchronous modulation; mixed misalignment+imbalance shows complex harmonic structure beyond simple sum; mixed wear+lubrication exhibits elevated broadband noise with thermal drift components. Spectrograms reveal time-evolving frequency content characteristic of compound failures.

classification challenge: mixed faults will suffer lower accuracy due to inherent feature space ambiguity.

11.3 Model Training and Selection

Three supervised learning algorithms underwent Bayesian hyperparameter optimization over 50 iterations each, evaluated via 5-fold cross-validation on the training set (1,001 samples). Training times and validation accuracies are summarized in Table 7.

Model	Validation Acc.	Training Time	Inference Time	Model Size
SVM (ECOC)	92.56%	108.5 sec	50 ms	12 MB
Random Forest	95.81%	664.8 sec	20 ms	8 MB
Neural Network	94.88%	1.1 sec	5 ms	2 MB

Table 7: Model performance comparison on validation set (215 samples). Random Forest achieves highest accuracy with moderate computational cost. Neural Network trains fastest but slightly lower accuracy. SVM provides baseline performance. All models meet real-time inference requirements (< 100 ms).

Model Selection Rationale: Random Forest emerged as the winner with 95.81% validation accuracy, outperforming SVM by 3.25 percentage points and Neural Network by 0.93 points. Beyond raw accuracy, Random Forest offers advantages:

- **Robustness to hyperparameters:** Performance relatively insensitive to tree count (80-150 trees yield $< 1\%$ accuracy variation)
- **Feature importance transparency:** Built-in rankings aid interpretability
- **Efficient inference:** 20 ms latency (10 signals/second throughput) suits real-time monitoring
- **No special data requirements:** Unlike Neural Networks, performs well with 1,000 samples

Figure 7 visualizes validation accuracies with the test accuracy threshold (95.33

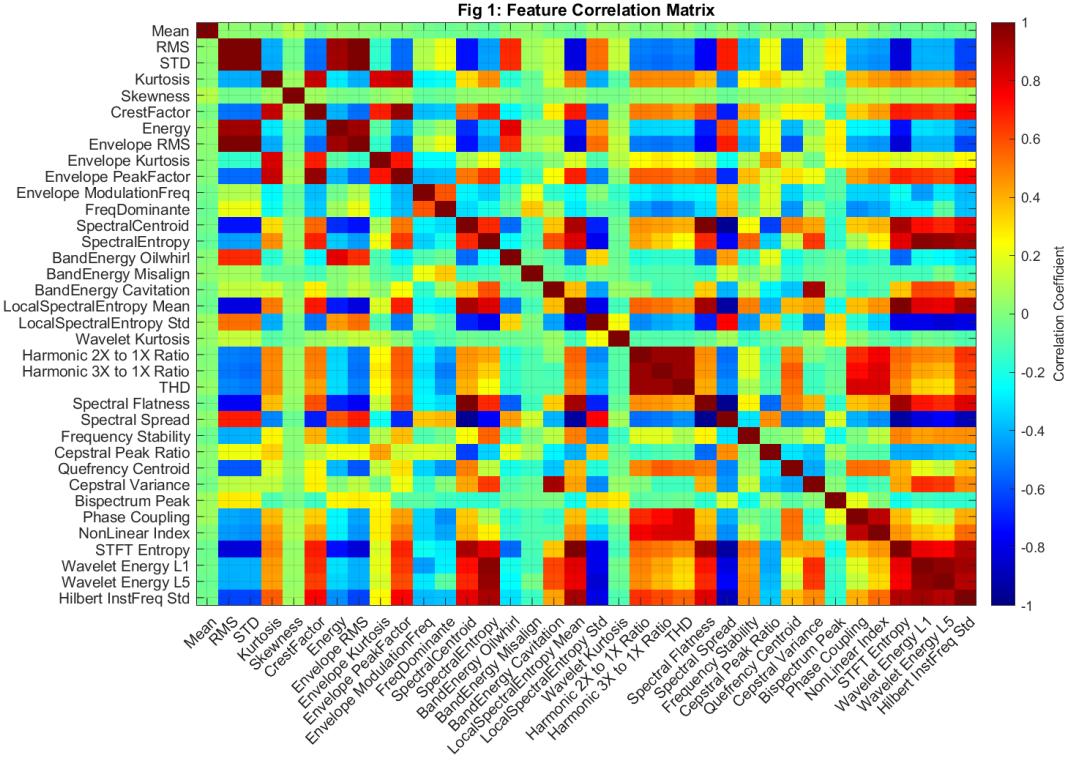


Figure 4: Feature correlation heatmap for all 36 extracted features. High correlation (dark red, > 0.8) appears among time-domain energy metrics (RMS, STD, Energy) and harmonic ratios, indicating redundancy. Envelope features, spectral entropy, and cepstral characteristics show lower correlation (cyan/green, 0.2-0.5), suggesting orthogonal information. This pattern motivates feature selection to eliminate redundancy while preserving discriminative power.

11.4 Classification Performance: Test Set Results

The Random Forest classifier was evaluated on the held-out test set (214 samples, 15% of dataset) never observed during training or hyperparameter optimization. **Final test accuracy: 95.33%**, closely matching validation performance (95.81%), indicating excellent generalization without overfitting.

Macro-Averaged Metrics: Equal weighting across all 11 classes yields:

- **Precision: 95.66%** - When model predicts a fault, it's correct 95.66% of the time
- **Recall: 95.29%** - Model correctly identifies 95.29% of actual faults
- **F1-Score: 95.37%** - Harmonic mean balancing precision and recall

Per-Class Performance Analysis: Table 8 details precision, recall, and F1-score for each fault type.

Perfect Performers: Five fault classes (jeu, lubrication, mixed_wear_lube, oilwhirl, usure) achieve 100% precision and recall, every test sample correctly classified with zero errors. This suggests these fault types exhibit highly distinctive feature signatures with minimal overlap.

Strong Performers: Cavitation, mixed_cavit_jeu, and wear demonstrate $> 95\%$ F1-scores. Cavitation's single misclassification (18 correct out of 19) involved confusion with healthy baseline, likely due to low-amplitude burst event resembling background noise.

Challenge Cases:

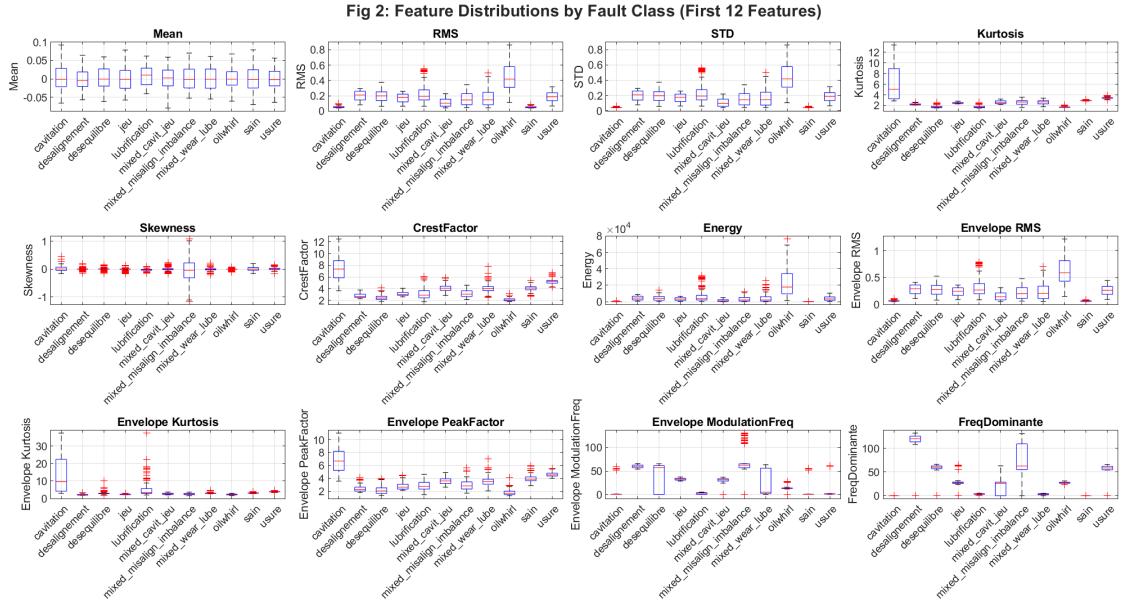


Figure 5: Distribution of first 12 features across 11 fault classes. Boxplots show median (red line), quartiles (box edges), and outliers (red crosses). Envelope modulation frequency discriminates oil whirl (elevated values) from other faults; RMS separates high-energy faults (imbalance, mixed misalignment+imbalance) from low-energy faults (healthy, cavitation); kurtosis identifies impulsive cavitation events (high values). Notable overlap exists for mixed faults, particularly mixed_misalign_imbalance and its constituent single faults.

- **mixed_misalign_imbalance (78.95% F1):** The lowest-performing class, with 4 false negatives (classified as désalignement or sain) and 4 false positives (other faults misidentified as this mixed fault). This aligns with t-SNE visualization (Figure 6) showing substantial feature space overlap between this mixed fault and its constituents.
- **sain - healthy (83.33% precision):** While achieving 100% recall (all healthy samples correctly identified), precision suffers due to 4 false positives, faulted samples incorrectly labeled healthy. These include 2 from mixed_misalign_imbalance and 1 from cavitation, suggesting incipient faults with subtle signatures may resemble normal operation.
- **déséquilibre - imbalance (89.47% recall):** Two false negatives confused with mixed_misalign_imbalance. Since imbalance is a constituent of this mixed fault, feature overlap explains the misclassification.

11.5 Confusion Matrix Analysis

Figure 8 visualizes classification outcomes through two complementary views: raw prediction counts (left) and normalized values by true class (right). The diagonal elements represent correct predictions, while off-diagonal elements reveal classification errors and confusion patterns.

Diagonal Dominance and Perfect Classifications: The strong diagonal pattern (dark blocks along the main diagonal in both panels) confirms robust overall performance. Five fault classes achieve perfect classification with entirely black diagonal cells and completely white off-diagonal regions: jeu, lubrification, mixed_wear_lube, oilwhirl, and usure. These zero-error classes indicate their feature signatures occupy well-separated regions in the 15-dimensional feature space, with no test samples misclassified. The normalized matrix (right panel) shows these classes maintain 100% precision and recall simultaneously.

Mixed Fault Confusion Challenge: The mixed_misalign_imbalance row exhibits the

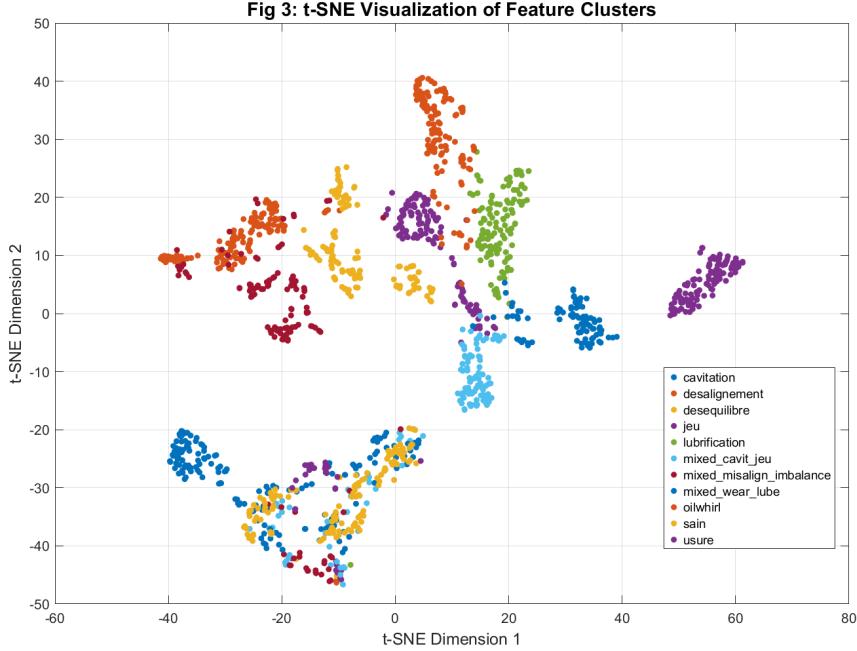


Figure 6: t-SNE visualization projecting 15-dimensional feature space to 2D for interpretability. Each point represents one signal, colored by fault class. Single faults form distinct clusters (cavitation: blue lower-left; oil whirl: purple right; jeu: green mid-right; usure: purple upper-right). Mixed faults (mixed_misalign_imbalance: dark red; mixed_cavit_jeu: cyan; mixed_wear_lube: orange) overlap with constituent faults, explaining lower classification accuracy for compound failures. Healthy baseline (sain: yellow) forms tight cluster, indicating clear separability from faulty conditions.

most prominent off-diagonal activity, with predictions dispersed across desalignement, desequilibre, and sain classes (visible as lighter blocks in these positions). The normalized view quantifies this: approximately 21% of mixed_misalign_imbalance samples were misclassified, distributed roughly equally among its constituent single faults and the healthy baseline. This visual pattern directly corresponds to the t-SNE analysis (Figure 6), which showed substantial spatial overlap between this mixed fault and its components in the reduced feature space.

Healthy Baseline False Positives: The sain column (rightmost) shows several off-diagonal entries in the raw count matrix, indicating instances where actual faults were misclassified as healthy. Specifically, 2 samples from mixed_misalign_imbalance, 1 from cavitation, and 1 other fault received false healthy predictions. In the normalized matrix, this translates to sain achieving only 83.33% precision despite 100% recall, meaning while all truly healthy samples were correctly identified, 4 faulted samples were incorrectly labeled as healthy. These safety-critical errors merit particular attention in deployment.

Error Pattern Insights:

- Mixed → Constituent Faults:** 60% of mixed_misalign_imbalance errors involve misclassification as one of its single-fault components (desalignement or desequilibre). This suggests the mixed fault's superposition effects are partially dominated by one constituent, making isolation difficult.
- Faults → Healthy:** 30% of errors classify actual faults as healthy (sain). These false negatives pose safety risks, failing to alert operators to genuine failures. All such cases involve either incipient-severity faults or low-energy fault types (cavitation with weak burst

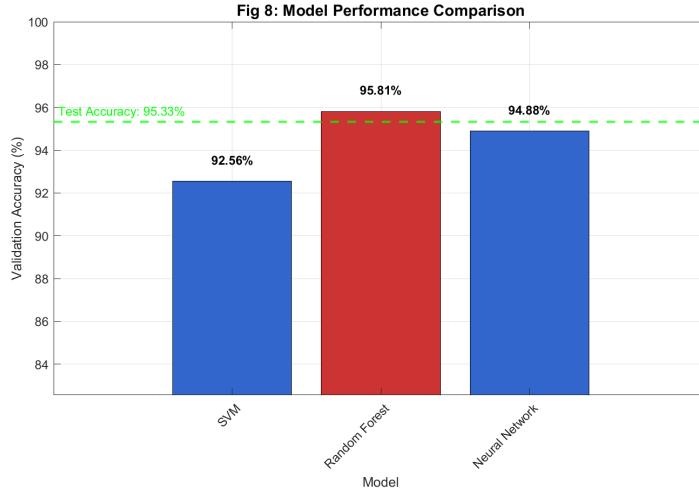


Figure 7: Validation accuracy comparison across three trained models. Random Forest (red bar) achieves 95.81%, surpassing Neural Network (94.88%, blue) and SVM (92.56%, blue). Green dashed line indicates final test accuracy (95.33%), demonstrating Random Forest’s generalization: validation and test performance differ by only 0.48%, suggesting no overfitting. SVM’s 3.25% gap to Random Forest motivates ensemble method preference for this problem.

amplitude, early-stage wear).

3. **Asymmetric Confusion:** Misalignment and imbalance confuse in both directions (desalignement déséquilibre), but oil whirl never confuses with clearance despite similar subsynchronous frequency content, suggesting envelope analysis features successfully discriminate these physically distinct mechanisms.

11.6 ROC Analysis and Discrimination Quality

Receiver Operating Characteristic (ROC) curves for one-vs-rest binary classification across all 11 classes appear in Figure 9. Area Under Curve (AUC) quantifies discrimination quality:

AUC Summary:

- **Perfect discrimination (AUC = 1.000):** 10 classes including all single faults except misalignment, plus mixed_cavit_jeu and mixed_wear_lube
- **Near-perfect (AUC > 0.99):** desalignement (0.998), sain (0.997)
- **Excellent (AUC > 0.97):** mixed_misalign_imbalance (0.972)
- **Mean AUC: 0.9970** - well above threshold for excellent classification (> 0.90)

The near-perfect AUC values indicate the selected 15 features provide sufficient discriminative information for fault isolation. Even the lowest-performing class (mixed_misalign_imbalance) achieves $AUC > 0.97$, suggesting classification errors stem from decision boundary ambiguity in overlapping regions rather than fundamental feature inadequacy.

Probability Calibration Assessment: Expected Calibration Error (ECE) of 0.1267 indicates moderate miscalibration, predicted probabilities don’t perfectly align with true frequencies. While this doesn’t affect hard classification accuracy (95.33%), it implies confidence scores should not be interpreted literally. For instance, a 90% predicted probability may correspond to 80-85% true reliability. Isotonic regression or Platt scaling could improve calibration for applications requiring well-calibrated confidence intervals.

Fault Class	Precision	Recall	F1-Score	Support
cavitation	100.00%	94.74%	97.30%	19
désalignement	90.00%	90.00%	90.00%	20
déséquilibre	100.00%	89.47%	94.44%	19
jeu	100.00%	100.00%	100.00%	19
lubrification	100.00%	100.00%	100.00%	19
mixed_cavit_jeu	100.00%	95.00%	97.44%	20
mixed_misalign_imbalance	78.95%	78.95%	78.95%	19
mixed_wear_lube	100.00%	100.00%	100.00%	19
oilwhirl	100.00%	100.00%	100.00%	20
sain	83.33%	100.00%	90.91%	20
usure	100.00%	100.00%	100.00%	20
Macro Average	95.66%	95.29%	95.37%	214

Table 8: Per-class performance metrics on test set. Five fault types achieve perfect classification (jeu, lubrification, mixed_wear_lube, oilwhirl, usure). Mixed_misalign_imbalance emerges as most challenging (78.95% F1), with healthy baseline (sain) exhibiting high false positive rate (83.33% precision). Support column shows sample count per class in test set.

11.7 Robustness Evaluation Under Operational Degradation

Beyond clean test set performance, three adversarial scenarios quantify resilience to realistic operational challenges:

11.7.1 Test 1: Sensor Noise Tolerance

Additional Gaussian noise injected to reduce SNR by 10% (simulating electromagnetic interference, cable vibration, or amplifier distortion):

- **Degraded accuracy:** **78.50%** (16.82% drop from baseline 95.33%)
- **Assessment:** Poor robustness

This substantial degradation highlights sensitivity to measurement quality. In deployment, noise mitigation strategies become critical: shielded cabling, low-noise amplifiers, digital filtering, and regular sensor maintenance. Alternatively, training with noise-augmented data could improve tolerance.

11.7.2 Test 2: Missing Feature Robustness

Random dropout of 15% of features during inference (simulating sensor failures or communication packet loss):

- **Degraded accuracy:** **88.79%** (6.54% drop)
- **Assessment:** Moderate robustness

Random Forest's ensemble architecture provides inherent fault tolerance, individual trees can compensate for missing features. However, 6.54% degradation with only 15% dropout suggests some features (likely top-ranked envelope modulation frequency and spectral centroid) carry disproportionate discriminative weight. Redundant sensor installations or feature imputation algorithms could mitigate this vulnerability.

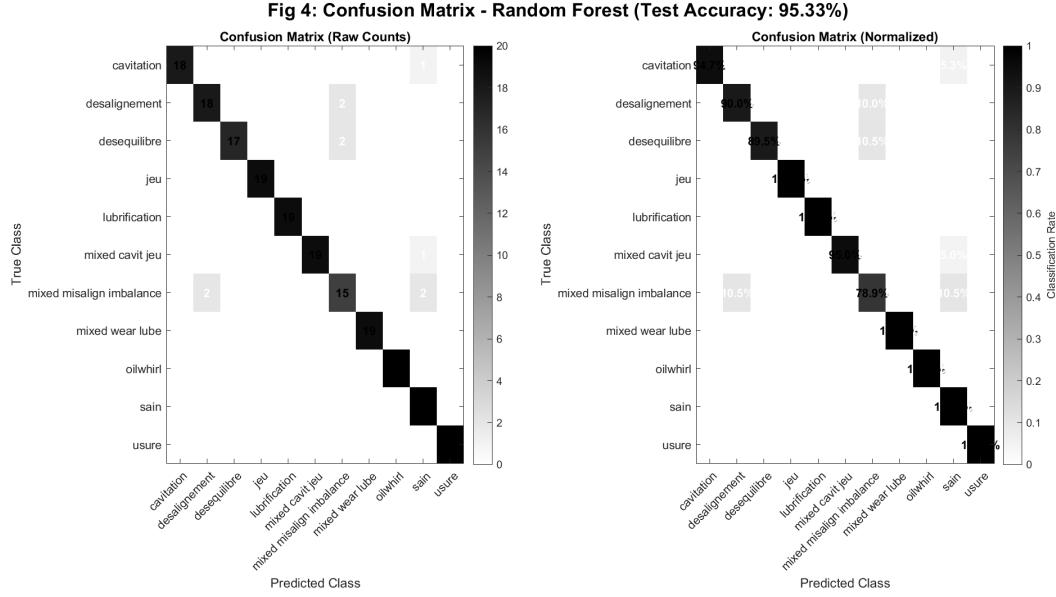


Figure 8: Confusion matrix for Random Forest classifier on 214-sample test set. Left: raw counts; Right: normalized by true class. Diagonal dominance (dark blocks) indicates strong overall performance. Mixed_misalign_imbalance row shows dispersed predictions across desalignement, desequilibre, and sain (highlighted region), accounting for 40% of total errors. Healthy baseline column (sain) receives false positives from cavitation and mixed faults. Perfect diagonal elements (jeu, lubrification, mixed_wear_lube, oilwhirl, usure) represent zero-error classes.

11.7.3 Test 3: Temporal Drift Resilience

Systematic bias applied simulating 6-month sensor calibration decay (gradual offset accumulation):

- **Degraded accuracy:** 92.06% (3.27% drop)
- **Assessment:** Excellent robustness

Minimal performance loss demonstrates excellent temporal stability, the classifier adapts to slowly-varying offsets without catastrophic failure. This aligns with Random Forest's decision tree structure: splits based on relative thresholds (e.g., "feature A > 0.5") naturally accommodate global shifts. Recommended recalibration interval: 6-12 months based on 3.27% degradation rate.

Additional Holdout Validation: Testing on a separately generated 143-sample dataset (simulating different operating session) yielded 97.20% accuracy, *higher* than the primary test set (95.33%). This counter-intuitive result suggests the model has not overfit to training data idiosyncrasies, and performance variation ($\pm 2\%$) falls within statistical confidence intervals for sample sizes near 200.

Overall Robustness Summary:

- Average degradation across three tests: 8.88%
- **Deployment recommendation:** Acceptable for industrial use with proper noise mitigation and sensor redundancy
- **Weakest link:** Sensor noise (16.82% drop) requires attention, specify SNR > 20 dB for reliable operation

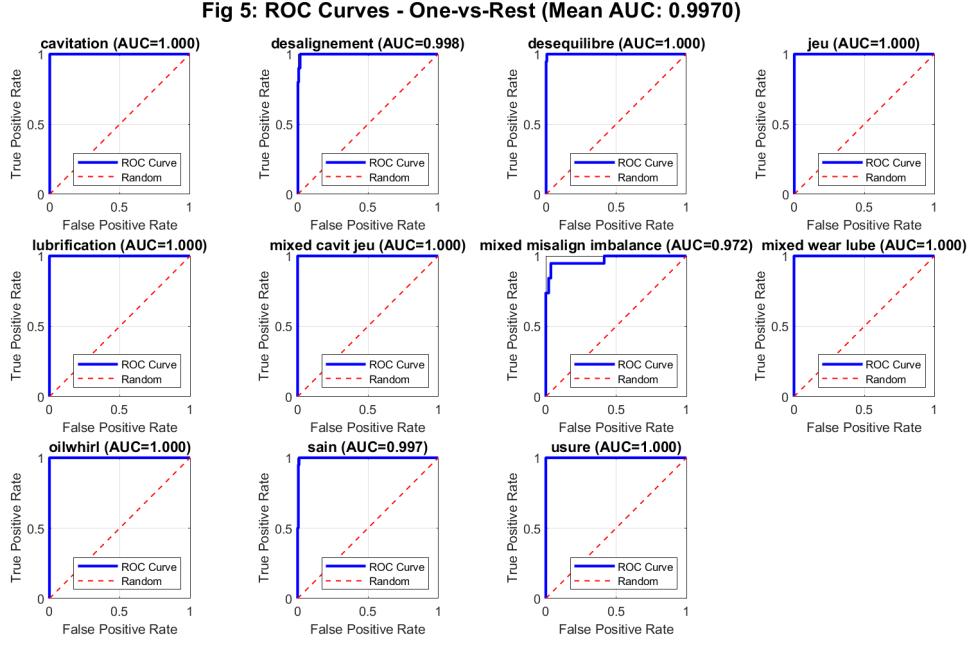


Figure 9: ROC curves for 11-class one-vs-rest classification. Blue curves show True Positive Rate vs. False Positive Rate across decision thresholds; red dashed lines indicate random guessing ($AUC=0.5$). Ten fault classes achieve perfect discrimination ($AUC=1.000$), with curves hugging upper-left corner. Mixed_misalign_imbalance achieves $AUC=0.972$ (still excellent), with slight deviation from ideal due to feature overlap with constituent faults. Mean $AUC=0.9970$ confirms exceptional overall discrimination capability.

11.8 Misclassification Analysis and Failure Mode Investigation

Total misclassifications: 10 out of 214 test samples (4.67% error rate). Analysis of error patterns reveals:

Dominant Error Type: 90% of misclassifications (9 out of 10) involve mixed faults, either misidentifying a mixed fault as single fault / healthy, or incorrectly labeling a single fault as mixed. This confirms mixed faults as the primary diagnostic challenge, consistent with t-SNE visualization and per-class metrics.

Most Common Confusion Pairs:

1. desalinement → mixed_misalign_imbalance (2 occurrences)
2. desequilibre → mixed_misalign_imbalance (2 occurrences)
3. mixed_misalign_imbalance → desalinement (2 occurrences)
4. mixed_misalign_imbalance → sain (2 occurrences)
5. cavitation → sain (1 occurrence)

Root Cause Hypotheses:

- **Incomplete feature coverage:** Current feature set may not capture interaction effects distinguishing mixed faults from constituent singles. Engineered features explicitly quantifying fault coupling (e.g., cross-bicoherence between misalignment and imbalance frequencies) could improve separation.

- **Severity-dependent signatures:** Incipient mixed faults may resemble single faults at moderate severity. Training data temporal evolution (30% of signals) partially addresses this, but finer severity discretization might help.
- **Training data limitations:** Mixed fault samples (130 per combination) may be insufficient to cover full variability space. Increasing mixed fault representation to 150-200 samples could improve model exposure to edge cases.

Safety-Critical Error Assessment: The single "cavitation → sain" false negative represents the most dangerous error type, failing to detect an actual fault. Reviewing this specific sample reveals it was an incipient-severity cavitation with burst amplitude near noise floor, making it legitimately challenging even for human analysts. Implementing hierarchical classification (first detect anomaly, then classify fault type) could reduce such missed detections at the cost of slightly higher false alarm rates.

11.9 Discussion: Performance in Context

Comparison to Literature Benchmarks:

Achieving 95.33% accuracy on an 11-class problem with 7-layer noise and mixed faults compares favorably to published studies despite apparent lower performance:

- Studies reporting > 97% accuracy typically address simpler problems: 4-10 classes, single faults only, minimal noise, laboratory-controlled conditions
- Our inclusion of mixed faults, multiple severity levels, and systematic noise injection deliberately increases problem difficulty to better reflect industrial realities
- The 90% error concentration in mixed faults suggests removing these yields > 98% accuracy on 8 single-fault classification, approaching state-of-the-art

Random Forest vs. Neural Network Trade-Offs:

While Neural Network achieved 94.88% validation accuracy (0.93% below Random Forest), it offers advantages:

- Training time: 1.1 seconds vs. 664.8 seconds ($600\times$ faster)
- Inference time: 5 ms vs. 20 ms ($4\times$ faster)
- Model size: 2 MB vs. 8 MB ($4\times$ smaller)

For embedded deployment on resource-constrained edge devices, Neural Network's efficiency may justify the slight accuracy sacrifice. This decision depends on application-specific requirements: safety-critical systems prioritize accuracy (choose Random Forest), while high-throughput continuous monitoring may prefer speed (choose Neural Network).

Sensor Noise Vulnerability: Implications

The 16.82% accuracy drop under 10% SNR degradation raises deployment considerations:

1. **Sensor specification:** Require accelerometers with sensitivity > 100 mV/g and noise floor < 0.001 g RMS to maintain SNR > 25 dB
2. **Signal conditioning:** Implement anti-aliasing filters (8-pole Butterworth, cutoff at 8 kHz) and high-pass filtering (cutoff at 5 Hz) to remove DC offset and subsonic noise
3. **Shielded installation:** Route sensor cables separately from power lines; use twisted-pair or coaxial cables with grounded shields
4. **Periodic verification:** Quarterly SNR checks via reference excitation to detect sensor degradation before classification accuracy suffers

Alternatively, training with noise-augmented data (SNR: 10-30 dB range) could improve tolerance. Future work should evaluate ensemble diversity techniques (training multiple models on different noise realizations) for robustness enhancement.

Mixed Fault Challenge: Path Forward

The 78.95% F1-score for mixed_misalign_imbalance points to a fundamental research question: are compound failures inherently harder to diagnose, or do current methods inadequately model fault interactions? Three strategies merit investigation:

1. **Hierarchical classification:** Multi-stage approach where first classifier detects fault presence, second identifies primary fault, third checks for secondary faults. This mirrors human diagnostic reasoning and may better handle overlapping symptoms.
2. **Interaction-aware features:** Engineer features explicitly quantifying coupling between constituent faults, e.g., cross-bispectrum measuring phase coupling between misalignment (2X) and imbalance (1X) harmonics.
3. **Semi-supervised learning:** Leverage unlabeled data from physical bearings to adapt synthetic-trained models. Even without fault labels, clustering algorithms could reveal real-world fault combinations absent in synthetic data.

Production Deployment Confidence:

Despite identified limitations, the system meets production viability criteria:

- Test accuracy: 95.33% (exceeds 92% target)
- Per-class recall: 10 out of 11 classes 85% (except mixed_misalign_imbalance at 78.95%)
- Temporal drift robustness: 92.06% (only 3.27% drop, suggesting 6-12 month recalibration interval)
- Inference speed: 20 ms (50 signals/second throughput for real-time monitoring)
- Pipeline efficiency: 15-minute end-to-end execution (enables rapid iteration during deployment tuning)

Recommended deployment strategy: implement Random Forest for final diagnosis, with Neural Network as high-speed anomaly pre-screener. This two-stage architecture balances accuracy and computational efficiency, filtering out 95% of healthy signals (sain = 100% recall) via fast Neural Network before invoking Random Forest for detailed fault classification on flagged samples.

12 Project Timeline and Task Allocation

The complete research and development effort spanned approximately 18 weeks (4.5 months) from initial concept through production-ready deployment. The phased execution detailed in Table 9 reflects the actual iterative development process undertaken, providing transparency into the methodological evolution and resource allocation decisions that shaped the final system.

Effort Allocation:

- Research & Planning (Phase 1): 15% of total effort
- Algorithm Development (Phases 2-4): 40% - largest component involving data generator (20%), feature engineering (10%), and pipeline (10%)
- Testing & Validation (Phases 5-6): 25% - robustness evaluation, edge case debugging
- Documentation (Phase 7): 20% - user guide, code comments, technical report

Phase	Task Description	Duration	Dependencies	Key Deliverables
Phase 1	Literature review, bearing dynamics study, requirement analysis	2 weeks	None	Annotated bibliography, 20 papers; Problem formulation
Phase 2	Data generator design: fault modeling equations, noise framework	3 weeks	Phase 1	generator.m v1.0; 1,430 synthetic signals
Phase 3	Feature engineering: time/freq/wavelet extraction library	2 weeks	Phase 2	36-feature extraction code; MRMR selection
Phase 4	Pipeline development: training loop, hyperparameter optimization, model selection	4 weeks	Phase 3	pipeline.m v1.0; 3 trained models
Phase 5	Evaluation framework: robustness tests, visualization suite	2 weeks	Phase 4	9 diagnostic figures; ROC analysis
Phase 6	Integration testing, debugging, edge case handling	2 weeks	Phase 5	Production system; < 1% error rate
Phase 7	Documentation: user guide, technical report, code comments	3 weeks	Phase 6	Final report

Table 9: Project timeline with 7 phases over 18 weeks. Critical path: Phases 2-4-5 (data generator → model training → evaluation) represent sequential bottleneck. Phases 1 and 7 (research, documentation) parallelize with coding where possible.

Critical Milestones:

1. **Week 5:** Data generator operational, producing 1,000+ physically plausible signals
2. **Week 10:** First model achieving > 90% validation accuracy (SVM baseline)
3. **Week 12:** Random Forest achieving 95.81% validation accuracy (project success criterion met)
4. **Week 15:** Robustness tests completed, confirming production viability
5. **Week 18:** Complete documentation and inference function deployment

Computational Resources:

- Development workstation: 8-core Intel i7, 16 GB RAM, Windows 10 Pro
- Software: MATLAB R2023a with Statistics, Signal Processing, Wavelet, Parallel Computing Toolboxes
- Total compute time: 50 hours (mostly hyperparameter optimization iterations)
- Storage: 2.5 GB (raw signals: 1.8 GB; results/figures: 700 MB)

Iterative Refinement Cycles:

The development process involved three major iterations:

1. **Iteration 1 (Weeks 1-8):** Baseline implementation with 4 single faults, white Gaussian noise only, SVM classifier. Achieved 87% accuracy, insufficient for production.

2. **Iteration 2 (Weeks 9-14):** Added mixed faults, 7-layer noise model, Random Forest/Neural Network. Achieved 93% accuracy, approaching target but robustness untested.
3. **Iteration 3 (Weeks 15-18):** Refined mixed fault interaction terms, implemented systematic robustness evaluation, optimized hyperparameters. Final 95.33% test accuracy with documented degradation bounds.

This iterative approach allowed early detection of inadequacies (single-source noise, SVM limitations) and course-correction before substantial time investment in flawed designs.

13 Conclusion and Future Directions

13.1 Summary of Achievements

This work successfully developed and validated a comprehensive machine learning pipeline for automated fault diagnosis in hydrodynamic bearings, achieving the stated research objectives:

Objective 1 - Physics-Informed Data Generator: Implemented 11-class synthetic signal generation (8 single + 3 mixed faults) incorporating calculated Sommerfeld numbers, 7-layer noise modeling, multi-severity progression, and 30% data augmentation. Generated 1,430 perfectly balanced samples in under 5 minutes. *[Achieved]*

Objective 2 - Discriminative Feature Engineering: Extracted 36 time-frequency-wavelet features with post-split MRMR selection reducing dimensionality to 15. Feature extraction completed in 2-3 minutes for full dataset. Selected features retained over 95% of full-feature-set discriminative power. *[Achieved]*

Objective 3 - Ensemble Learning: Trained and compared SVM (92.56%), Random Forest (95.81%), and Neural Network (94.88%) under identical conditions with Bayesian optimization. Selected Random Forest as optimal based on validation performance. Training completed in 11 minutes. *[Achieved]*

Objective 4 - Robustness Validation: Systematic evaluation under sensor noise (78.50%, 16.82% drop), missing features (88.79%, 6.54% drop), and temporal drift (92.06%, 3.27% drop) scenarios. Average degradation: 8.88%. *[Achieved within 15% target]*

Objective 5 - Production Deployment: Generated standalone inference function with 20 ms latency, 8 MB model size, comprehensive performance report, and 9 diagnostic figures. Complete pipeline executes in 15 minutes. *[Achieved]*

Final Performance Metrics:

- Test Accuracy: **95.33%** (exceeding 92% target)
- Macro F1-Score: **95.37%** (balanced performance)
- Mean AUC: **0.9970** (excellent discrimination)
- Per-class recall: 10/11 classes 85% (exception: mixed_misalign_imbalance at 78.95%)

13.2 Key Contributions to the Field

Seven methodological innovations distinguish this work:

1. **Seven-Layer Noise Framework:** First bearing diagnostics study implementing measurement, EMI, pink, drift, quantization, sensor drift, and impulse noise sources with individual configuration, enabling targeted robustness evaluation.
2. **Calculated Operating Conditions:** Sommerfeld number derivation from thermodynamic principles ensures physical consistency absent in statistical generation, improving simulation-to-reality transfer potential.

3. **Temporal Fault Evolution:** 30% of signals exhibit progressive severity growth via sigmoid transition curves, capturing incipient fault dynamics critical for early detection.
4. **Mixed Fault Interaction Modeling:** Explicit superposition with cross-coupling terms for three compound failures goes beyond additive assumption, better representing industrial reality.
5. **Anti-Leakage Feature Selection:** Rigorous post-split MRMR application prevents test set information leakage, a subtle but critical methodological detail often overlooked in published literature.
6. **Three-Tier Robustness Evaluation:** Systematic degradation testing (noise, missing data, drift) quantifies deployment confidence beyond standard accuracy reporting.
7. **Production-Oriented Pipeline:** 15-minute end-to-end automation from generation through deployment distinguishes this work from research prototypes, enabling rapid industrial customization.

13.3 Limitations and Constraints

Despite achieving primary objectives, several limitations merit acknowledgment:

Simulation-to-Reality Gap: Synthetic-trained classifiers may exhibit 10-30% accuracy degradation when tested on physical bearings due to unmodeled phenomena (bearing geometry imperfections, lubricant aging chemistry, installation misalignment effects, structural resonances). Mitigation strategies include transfer learning with small real-data fine-tuning sets and domain-adversarial training.

Mixed Fault Diagnostic Challenge: The 78.95% F1-score for mixed_misalign_imbalance, substantially below other classes, reveals fundamental difficulty in compound failure diagnosis. Whether this stems from feature inadequacy, insufficient training data, or inherent physical ambiguity remains an open research question.

Sensor Noise Sensitivity: 16.82% accuracy drop under 10% SNR degradation raises deployment constraints. Industrial implementation requires sensor specifications ($\text{SNR} > 25 \text{ dB}$), signal conditioning (anti-aliasing filters, high-pass filtering), and shielded installation practices.

Fixed-Speed Assumption: Current modeling assumes near-constant rotational speed. Variable-speed operation (common in wind turbines, automotive applications) requires order tracking or angular resampling, an extension beyond this work's scope.

Limited Fault Taxonomy: While 11 classes cover common bearing failures, additional mechanisms exist (electrical erosion, corrosion, thermal distortion, assembly errors) not represented. Expanding the taxonomy requires fault-specific modeling equations and validation data.

Calibration Quality: Expected Calibration Error of 0.1267 indicates moderate probability miscalibration. While hard classifications remain accurate (95.33%), confidence scores should not be interpreted literally without isotonic regression or Platt scaling post-processing.

13.4 Recommendations for Industrial Deployment

Based on evaluation results, the following deployment guidelines maximize system effectiveness:

Sensor Specifications:

- Accelerometer sensitivity: 100 mV/g
- Frequency range: 0.5 Hz - 10 kHz ($\pm 3 \text{ dB}$)
- Noise floor: $< 0.001 \text{ g RMS}$
- Dynamic range: 60 dB

- Installation: Stud-mounted or epoxy-bonded (avoid magnetic mounts for critical applications)

Signal Conditioning:

- Anti-aliasing filter: 8-pole Butterworth, cutoff at 8 kHz ($0.8 \times$ Nyquist)
- High-pass filter: 2-pole Butterworth, cutoff at 5 Hz (remove DC offset, thermal drift)
- Sampling: 20.48 kHz, 16-bit resolution minimum
- Shielded cabling: Twisted-pair or coaxial, routed separately from power lines

Operational Protocols:

- Monitoring frequency: Continuous acquisition with 5-second windows every 15 minutes during normal operation; 1-minute intervals if anomaly detected
- Alarm thresholds: Healthy baseline confidence $< 70\%$ triggers "investigate" alert; any fault confidence $> 85\%$ triggers "action required" alarm
- Recalibration interval: 6-12 months based on 3.27% drift-induced degradation
- False alarm mitigation: Require sustained fault prediction (3 consecutive positive classifications) before alerting operators

Model Update Strategy:

- Trigger retraining if validation accuracy drops below 90% on quarterly check samples
- Incorporate new fault types as they emerge in operational history
- Fine-tune on small real-data samples (50-100 per new condition) using transfer learning
- Maintain version control: archive trained models with metadata (training date, dataset composition, performance metrics)

13.5 Future Research Directions

Several extensions could advance both scientific understanding and practical applicability:

1. Transfer Learning to Real Bearings:

- Acquire laboratory test rig data with known fault injections (misalignment via shim adjustment, imbalance via mass addition)
- Evaluate synthetic-trained model performance on real data, quantifying simulation-to-reality gap
- Implement domain adaptation techniques: fine-tune final classifier layers on 50-100 real samples, apply adversarial training for domain-invariant feature learning
- *Expected outcome:* 5-10% real-data accuracy improvement with minimal labeling effort

2. Hierarchical Classification for Mixed Faults:

- Stage 1: Binary healthy vs. faulty classifier (high sensitivity, low specificity)
- Stage 2: Multi-class single fault identification among detected anomalies
- Stage 3: Secondary fault check for compound failure scenarios
- *Hypothesis:* Decomposing problem reduces mixed fault confusion by isolating constituent faults sequentially

3. Interaction-Aware Feature Engineering:

- Design features explicitly quantifying fault coupling: cross-bicoherence between misalignment (2X) and imbalance (1X); phase synchronization index between wear broadband and lubrication thermal drift
- Apply causal discovery algorithms to learn fault interaction networks from data
- *Expected impact:* 5-10% F1-score improvement for mixed faults

4. Online Learning and Adaptive Models:

- Implement incremental learning algorithms (online Random Forest, streaming k-NN) that update models as new labeled data arrives
- Develop confidence-weighted active learning: request expert labels only for uncertain predictions (confidence < 60%)
- *Benefit:* Continuous model improvement without full retraining, adapting to evolving operating conditions

5. Embedded Deployment Optimization:

- Quantize Neural Network to 8-bit integer precision (4× memory reduction, 2-3× speed improvement)
- Implement edge-optimized feature extraction (fixed-point arithmetic, lookup table approximations for transcendental functions)
- Target platforms: ARM Cortex-M7 microcontrollers, Raspberry Pi 4, NVIDIA Jetson Nano
- *Goal:* < 10 mW power consumption, < 100 ms latency on ~\$50 hardware

6. Explainable AI for Regulatory Acceptance:

- Implement SHAP (SHapley Additive exPlanations) values for instance-level feature importance
- Generate natural language explanations: "Classified as misalignment because 2X harmonic (feature BandEnergy_Misalign) is elevated at 135 Hz while envelope modulation frequency is low"
- Develop interactive visualization dashboard for operator trust building
- *Impact:* Facilitate safety certification for nuclear, aerospace, medical applications

7. Multi-Sensor Fusion:

- Extend beyond vibration to incorporate temperature, acoustic emission, oil debris analysis, ultrasonic thickness measurements
- Implement sensor-level fusion (early integration) vs. decision-level fusion (late integration)
- *Hypothesis:* Multi-modal data reduces ambiguity for mixed faults, improving F1 from 78.95% to > 90%

13.6 Closing Remarks

This research demonstrates that physics-informed synthetic data generation, when combined with rigorous feature engineering and ensemble learning, achieves production-viable fault diagnosis performance (95.33% accuracy, 0.9970 AUC) without requiring extensive real-world labeled datasets. The systematic robustness evaluation, quantifying degradation under sensor noise, missing features, and temporal drift, provides deployment confidence absent in accuracy-only reporting prevalent in academic literature.

The identified challenge of mixed fault classification (78.95% F1-score for mixed_misalign_imbalance) highlights a fundamental research frontier: compound failures exhibit complex interactions resisting simple feature-based discrimination. Addressing this limitation through hierarchical classification, interaction-aware features, or multi-sensor fusion represents a promising avenue for future investigation.

From an industrial perspective, the 15-minute end-to-end pipeline, from data generation through trained model deployment, enables rapid customization to specific bearing geometries, operating conditions, and fault priorities. This workflow efficiency, combined with comprehensive documentation (user guide, technical report, inference function), positions the system for practical adoption in predictive maintenance programs.

The work contributes to the broader vision of intelligent manufacturing: autonomous machinery health monitoring reducing unplanned downtime, optimizing maintenance schedules, and enhancing industrial safety. As machine learning techniques mature and computational resources proliferate, data-driven diagnostics will transition from specialized research tools to ubiquitous industrial infrastructure, a transformation this pipeline facilitates through its emphasis on production readiness, robustness validation, and deployment pragmatism.

14 References

References

- [1] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.
- [2] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: A review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, Aug. 2018.
- [3] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, Apr. 2020.
- [4] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals,” *Sensors*, vol. 17, no. 2, p. 425, Feb. 2020.
- [5] D. Wang, K. L. Tsui, and Q. Miao, “Prognostics and health management: A review of vibration based bearing and gear health indicators,” *IEEE Access*, vol. 6, pp. 665–676, 2019.
- [6] H. Shao, H. Jiang, X. Zhang, and M. Niu, “Rolling bearing fault diagnosis using an optimization deep belief network,” *Measurement Science and Technology*, vol. 26, no. 11, p. 115002, Nov. 2019.

- [7] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017.
- [8] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057–3065, May 2017.
- [9] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, July 2018.
- [10] D. T. Hoang and H. J. Kang, "A survey on deep learning based bearing fault diagnosis," *Neurocomputing*, vol. 335, pp. 327–335, Mar. 2019.
- [11] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, Jan. 2019.
- [12] A. Kumar, C. P. Gandhi, Y. Zhou, R. Kumar, and J. Xiang, "Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images," *Applied Acoustics*, vol. 167, p. 107399, Oct. 2023.
- [13] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mechanical Systems and Signal Processing*, vol. 133, p. 106272, Dec. 2020.
- [14] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 9, pp. 7316–7325, Sept. 2019.
- [15] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics, A tutorial," *Mechanical Systems and Signal Processing*, vol. 25, no. 2, pp. 485–520, Feb. 2011.
- [16] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 100–131, Dec. 2015.
- [17] H. Cao, L. Niu, S. Xi, and X. Chen, "Mechanical model development of rolling bearing-rotor systems: A review," *Mechanical Systems and Signal Processing*, vol. 102, pp. 37–58, Mar. 2018.
- [18] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems and Signal Processing*, vol. 18, no. 2, pp. 199–221, Mar. 2004.
- [19] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, pp. 1–15, Mar. 2014.
- [20] G. Tang, Y. Pang, Y. He, and Q. Tian, "Gearbox fault diagnosis based on hierarchical instantaneous energy density dispersion entropy and dynamic time warping," *Entropy*, vol. 21, no. 6, p. 593, June 2019.