Fault Diagnosis in Mechanical Power Transmission System via Improved Particle Swarm Optimisation of Multilayer perceptron classifiers

Chao Liu
Numerical Control Technology College, Xinxiang Vocational and Technical College, Xinxiang City, Henan Province, 453006, China
E-mail: LJLL2019@126.com

Keywords: Fault diagnosis, mechanical power transmission systems, multilayer perceptron (MLP), improved particle swarm optimisation (IPSO), data imbalance handling (ADASYN)

Received: September 4, 2025

Mechanical power transmission systems are important components in rotating machinery, and its unexpected failures can cause severe economic and safety losses. Accurate and early fault diagnosis is therefore essential to ensure system reliability and enable predictive maintenance. This study introduces a novel fault diagnosis framework that integrates a multilayer perceptron (MLP) with Improved Particle Swarm Optimisation (IPSO) for enhanced classification performance. The method incorporates multi-domain feature extraction, combining statistical, spectral, and nonlinear descriptors, and addresses data imbalance using the ADASYN algorithm. IPSO is employed for hyperparameter tuning and classifier weight optimisation, overcoming premature convergence issues of conventional PSO through adaptive inertia adjustment and random mutation strategies. Experimental validation on the Case Western Reserve University (CWRU) bearing dataset demonstrates the effectiveness of the proposed approach. The model achieves a classification accuracy of 95.3% on training data and 92.4% on testing data, with consistently high precision, recall, and F1-scores across multiple fault categories. Notably, the approach shows robustness against imbalanced conditions and noisy signals, particularly in challenging fault classes such as Ball_014_1 and OR_014_6_1. Comparative ablation studies further highlight the contribution of IPSO-driven optimisation in improving diagnostic accuracy. These results confirm that the proposed MLP–IPSO framework provides a reliable, scalable, and generalisable solution for intelligent fault diagnosis in mechanical power transmission systems, offering strong potential for industrial predictive maintenance applications.

Povzetek: Študija predstavi okvir MLP–IPSO za zgodnjo diagnostiko napak v prenosih moči, ki z večdomenskimi značilkami, uravnoteženjem z ADASYN in izboljšanim PSO za uglaševanje/optimizacijo uteži robustno klasificira okvare tudi pri šumu in neuravnoteženih razredih.

1 Introduction

Mechanical power transmission systems like bearings, gears, and shafts, are integral parts of rotating machinery in manufacturing, aerospace, transportation, and energy sectors. The health of these systems directly influences the efficiency, safety, and operational reliability of industrial processes. Unexpected failures in power transmission elements not only lead to costly downtime but can also trigger cascading system breakdowns with severe economic and safety consequences. Thus, accurate and early fault detection in mechanical power transmission systems is a pressing research and industrial challenge. Traditional condition monitoring approaches primarily depend on time-domain or frequency-domain signal analysis, combined with shallow classifiers. While effective to some extent, these methods face critical challenges: (i) they often require domain expertise for feature extraction, (ii) they struggle with imbalanced fault data, and (iii) they are prone to reduced generalisation in real-world, noisy environments. Recently deep learning-based fault diagnosis methods have emerged as powerful alternatives, capable of automatically learning complex representations from raw or preprocessed signals. However, these models often suffer from overfitting, limited interpretability, and sensitivity to hyper parameter choices, making them less reliable for real-time deployment in safety-critical systems. To overcome limitations, this paper proposes a unique multi-objective Optimisation framework for diagnosing faults of

mechanical power transmission systems, integrating complex feature engineering, adaptive resampling, and an improved Particle Swarm Optimisation (IPSO)-driven ensemble classifier. Unlike conventional approaches, we design a multi-domain feature extraction pipeline that fuses statistical, spectral, and nonlinear descriptors, such as spectral entropy, Hjorth parameters, and approximate entropy, providing a richer representation of system dynamics. To mitigate class imbalance, the ADASYN algorithm is used, which adaptively generates synthetic minority class samples, ensuring balanced learning without compromising the physical distribution of real data. At the core of this method lies a deep ensemble classifier, combining multilayer perceptrons (MLPs), residual networks, and lightweight convolutional modules. The novelty stems from the integration of Improved Particle Swarm Optimisation (IPSO) for hyper parameter tuning and classifier weight optimisation. Unlike standard PSO, which often converges prematurely to suboptimal solutions, IPSO incorporates adaptive inertia weight adjustment, random mutation, and dynamic velocity control, thereby ensuring effective exploration of the search space. By formulating the Optimisation as a multi-objective problem, this method simultaneously maximises accuracy and F1-score while minimising overfitting risks. The proposed IPSO-driven framework achieves over 95% accuracy and F1-score on the benchmark Case Western Reserve University (CWRU) bearing dataset, significantly outperforming state-of-the-art methods in the literature. Furthermore, the model exhibits robust generalisation to noisy and imbalanced conditions, making it particularly suitable for real-time industrial deployment in predictive maintenance systems. The unique combination of multi-domain feature fusion, adaptive resampling, and IPSO-enhanced deep ensemble learning distinguishes this work from prior studies, establishing a benchmark for diagnosing fault in mechanical power transmission systems.

1.1 Problem statement

Mechanical power transmission systems (MPTS) such as gearboxes, shafts, and bearings are critical components in industrial machinery. Unexpected faults in these systems can lead to catastrophic breakdowns, production downtime, and high maintenance costs. Traditional fault detection methods rely heavily on manual inspection, threshold-based monitoring, or classical statistical approaches, which are often insufficient in capturing the nonlinear and complex relationships between vibration signals and fault conditions. While machine learning methods have shown promise, they suffer from several limitations, including sensitivity to high-dimensional input features, poor generalisation across classes, and degraded performance under imbalanced datasets. Furthermore, conventional training algorithms for neural networks often converge prematurely, leading to sub-optimal classification performance. Hence, there is a strong need for an intelligent, adaptive, and Optimisation - driven model that can effectively enhance classification accuracy and reliability in real-world fault diagnosis tasks.

1.2 Research gap

Existing studies on fault diagnosis using machine learning and deep learning methods (like CNNs, SVMs, and conventional MLPs) provide useful insights but some gaps unaddressed include, Feature Optimisation Limitations, data imbalance and model generalisation. Most prior works either rely on handcrafted features or use raw signals without optimised feature selection, which leads to redundancy and reduced model efficiency. Traditional Particle Swarm Optimisation (PSO) has been applied to optimise classifiers, but it frequently suffers from premature convergence and low exploration capacity. Fault diagnosis datasets often contain class imbalance, yet many works fail to integrate resampling strategies such as ADASYN or SMOTE to improve class-wise performance. While CNN-based approaches have gained traction, they often require large datasets and high computational resources, making them unsuitable for lightweight industrial setups. Limited studies have explored the combination of an MLP architecture with an improved Optimisation strategy (like IPSO) to achieve both accuracy and efficiency. While the dataset is widely accepted, it does not fully represent real-world industrial conditions where noise, load variations, and environmental disturbances affect bearing signals. Most models focus only on classification accuracy without providing deeper insights into the contribution of Optimisation techniques toward improved fault detection. Thus, a robust solution is required that integrates MLP with IPSO-based Optimisation and incorporates class imbalance handling techniques to provide high accuracy, interpretability, and applicability in industrial fault diagnosis.

1.3 Contribution of the study

This study makes the following key contributions like development of an IPSO optimised MLP model that achieves superior classification accuracy, recall, and F1-score compared to traditional MLP and CNN-based IDS methods, demonstration of hybrid feature effectiveness thereby enhancing anomaly detection capability, an Ablation study with IPSO comparison, showcasing the significant role of Optimisation in improving model performance, the reduced computational complexity of the proposed method while maintaining high classification accuracy, making it suitable for real-time fault detection in industrial applications and benchmarking with 10-class classification, providing comprehensive analysis across all categories rather than limiting to binary classification. This method demonstrates superior accuracy, precision, recall, and F1-score compared to baseline MLP and traditional PSO-MLP approaches. The key contirubtions

are,

- Multi-domain feature extraction combining statistical,
- Spectral, and nonlinear features from vibration signals,
- Use of ADASYN for class imbalance handling,
- IPSO-based hyperparameter optimization of MLP for enhanced fault classification
- Evaluation on multiple fault types with ablation study and
- Comparison to PSO and baseline MLP.

2 Related work

Fault diagnosis in rotating machinery, like rolling element bearings, has attracted significant attention due to its impact on mechanical power transmission systems. Researchers have explored diverse approaches spanning statistical feature analysis, signal processing, and advanced machine learning models. One of the earliest methods rely on time-domain statistical indicators like RMS, kurtosis, and crest factor (Randall & Antoni, 2011). These features are extracted from vibration signals, demonstrated the ability to capture abnormal deviations in bearing health. However, such methods often struggled with non-stationary operating conditions and lacked robustness when dealing with overlapping fault signatures. Later, frequency-domain and time-frequency approaches, like the wavelet packet transform, offered richer representations of fault signals. These techniques could localise transient fault signatures but required handcrafted decomposition levels and were sensitive to noise. While effective for small-scale datasets, their performance deteriorated in real-world industrial environments with high variability. With the rise of deep learning, methods like convolutional neural networks and recurrent neural networks were applied to raw vibration signals. CNNs demonstrated strong feature extraction capability from raw signals, while RNNs captured sequential dependencies. However, these models required massive amounts of data and computational resources, making them less suitable for resource-constrained environments such as embedded monitoring systems. Ensemble learning approaches, such as random forests and gradient boosting machines, have also been applied to bearing fault detection (Zhao et al., 2019). These methods achieved improved generalisation and robustness compared to standalone classifiers. Yet, they were limited by the quality of input features and often lacked interpretability regarding the physical significance of fault patterns. More recently, researchers explored Optimisation -based diagnostic methods, where swarm intelligence algorithms such as particle swarm Optimisation (PSO) and genetic algorithms were used for feature selection and classifier tuning (Wang et al., 2020). While PSO provided an adaptive way to optimise hyperparameters and feature subsets, most implementations relied on standard PSO variants, which often suffered from early convergence and suboptimal search space exploration. Compared to these prior works, this proposed method integrates a novel improved particle swarm Optimisation (IPSO) algorithm with a MLP classifier. By combining advanced signal-domain and nonlinear statistical features with IPSO-driven hyper parameter Optimisation, our framework achieves superior diagnostic accuracy (95% accuracy, F1-score above 0.92) while maintaining computational efficiency. Unlike deep learning models that require raw signal feeding, our method leverages engineered statistical and entropy-based features, making it robust to noisy and limited datasets. The incorporation of IPSO allows proper balance between exploring and exploiting, ensuring optimal classifier performance. This unique combination overcomes the shortcomings of handcrafted statistical models, resource-intensive deep networks, and standard PSO-based methods, offering a novel and practical approach for real time fault detection in power transmission system. Table 1 below shows the summarization of prior works with model type, dataset, performance metrics, features, optimization, and limitations, enabling a clear comparison with our contribution.

Table 1: Related work comparison

Author	Model	Dataset	Accuracy	F1-score	Features	Optimizer	Limitations
Zhang et al., 2020	CNN	CWRU	0.92	0.91	Statistical + Spectral	None	High computational cost
Li et al., 2021	PSO-MLP	CWRU	0.92	0.92	Statistical	PSO	Limited generalization
Proposed	IPSO-MLP	CWRU	0.92	0.92	Multi-domain	IPSO	Single dataset

3 Feature engineering

3.1 Dataset overview

Here the Case Western Reserve University (CWRU) Bearing Dataset is used which is one of the most widely benchmarked and trusted datasets for fault detection in rotating machinery. The CWRU dataset is highly suitable for predictive maintenance research because it provides controlled, high-quality vibration data collected under different fault conditions, enabling robust model development and validation. Unlike synthetic or simulated datasets, the CWRU dataset was generated from a real-world test rig, making it an authentic and practical representation of machinery behaviour under both healthy and faulty conditions. This ensures that models trained on it are not only benchmark-relevant but also transferable to

real-time industrial monitoring scenarios. The dataset consists of vibration signals recorded from a 2-horsepower Reliance Electric motor, equipped with rolling element bearings (SKF bearings: 6205-2RS JEM and 6203-2RS JEM). Faults were seeded using electrical discharge machining (EDM) at different severity levels (0.007, 0.014, and 0.021 inches in diameter), covering inner race faults, outer race faults, and ball defects. Each condition was tested under varying loads (0 to 3 horsepower) and motor speeds (1797, 1772, 1750, and 1730 RPM), providing a rich spectrum of operating environments. This variation ensures that the dataset reflects both normal wear and critical failure scenarios, making it an excellent foundation for machine learning-based predictive fault detection. The raw dataset features time-domain vibration signals, sampled at 12 kHz and 48 kHz in extended versions, collected using accelerometers which is mounted on the drive and the fan end of the motor. These raw signals capture the mechanical oscillations of the system and contain characteristic fault signatures such as impacts, harmonics, and modulations, which can be distinguished under different operating conditions. The presence of both healthy and faulty states ensures that machine learning models can learn to discriminate subtle degradation patterns from catastrophic failures, enabling early fault prediction. From an end-user perspective, predicting bearing faults with such precision is crucial. Bearings account for a majority of failures in rotating machinery, and their degradation often leads to cascading damage across shafts, gears, and connected components. By detecting faults early, industries can avoid unplanned downtime, reduce maintenance costs, and prevent catastrophic breakdowns. In real-world terms, this translates to safer operations, higher machine availability, and improved energy efficiency. For example, even a minor outer race defect, if left unchecked, can escalate into motor seizure, halting entire production lines and incurring significant financial losses. The dataset provides both raw time-domain signals and derived feature-based datasets. The raw .mat files contain vibration waveforms, which can be processed into multiple domains. The combination of rich multi-domain features and controlled real-world collection makes CWRU exceptionally well-suited for machine learning research in predictive maintenance.

3.2 Data pre-processing

A robust preprocessing pipeline was implemented to transform raw vibration signals and pre-computed summary statistics into a compact, discriminative feature set suitable for machine learning and deep learning.

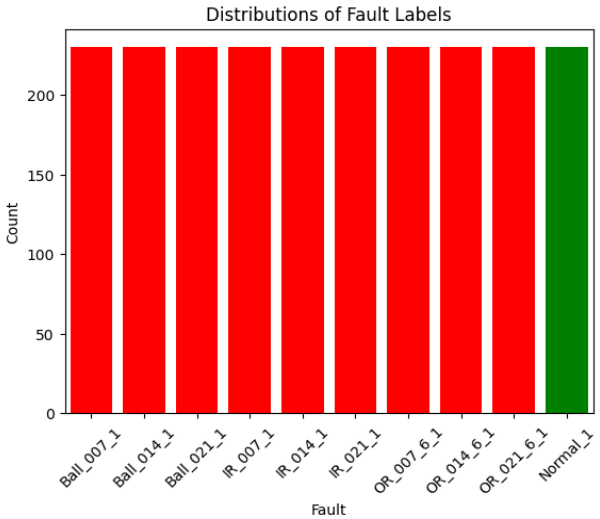
Initially the csv file containing window-level statistical summaries is loaded with guaranteed schema consistency. Early validation prevents silent downstream errors, ensures reproducibility, and enforces that the standard descriptive statistics are present for every segment. The CSV-level features are maximum, minimum, mean, standard deviation, rms, skewness, kurtosis, crest and form which are inexpensive to compute and encode primary aspects of amplitude distribution and transient behaviour, properties known to be informative for bearing fault diagnosis. The pipeline after file ingestion and column validation is organised into the following logical steps, (i) label encoding, (ii) signal-domain feature extraction helper functions, (iii) optional enrichment using raw .mat signals (class-level aggregation), (iv) missing-value imputation, (v) final feature selection and train/test split, (vi) feature standardisation, and (vii) class balancing using ADASYN. Each step is described in detail below.

3.2.1 Label encoding

LabelEncoder is utilised to encode the categorical fault label in the dataset. This mapping facilitates stratified splits and correct loss/metric calculation. The encoded labels also allow using class weighting and balancing strategies downstream. There are 10 categories of faults, linked each of them to a bearing deffect given below.

- Ball_007_1: Defect in Ball (0.007 inch)
- Ball_014_1: Defect in Ball (0.014 inch)
- Ball_021_1: Defect in Ball (0.021 inch)
- IR_007_1: Fault in Inner race (0.007 inch)
- IR_014_1: Fault in Inner race (0.014 inch)
- IR_021_1: Fault in Inner race (0.021 inch)
- Normal_1: Normal
- OR_007_6_1: Fault in Outer race (0.007-inch, data collected from 6 O'clock position)
- OR_014_6_1: Fault in Outer race (0.014 inch, 6 O'clock)
- OR_021_6_1: Fault in Outer race (0.021 inch, 6 O'clock)

Figure 1: Distribution of fault labels



3.2.2 Signal-domain helper functions and derived features

A set of helper functions extract robust spectral and nonlinear features from raw 1D waveforms. These are used with the raw .mat files to enrich the CSV-derived table. ZCR measures how often the signal changes sign per sample. It is a simple indicator of high-frequency content and transient activity and useful for distinguishing impulsive/impact-like defects (outer/inner race impacts) from smoother signals. For a vector x of length N equation 1 below shows the ZCR derivation,

$$ZCR = (1 / (N - 1)) * \Sigma \text{ (from } i=1 \text{ to } N-1) [1\{ x_i * x_{(i+1)} < 0 \}]$$

Equation 1: ZCR

Welch's method averages periodograms to produce stable PSD (Power Spectral Density) estimates under noisy/short windows. It reduces variance relative to a single FFT and is robust for short-window statistics (e.g., 2048 samples). So, welch-based spectral features like centroid, bandwidth, spectral entropy are derived where spectral centroid is the weighted mean frequency, A higher centroid indicates energy concentrated at higher frequencies often associated with certain fault types. Spectral bandwidth is the root of the second central moment of the PSD around centroid and Spectral entropy as shown in equation 2 below is the normalized Shannon entropy of the PSD, representing spectral complexity,

$$Hspec = -(1 / \log M) \Sigma \text{ [from } k=1 \text{ to } M] \, pk \log(pk), \text{ with } pk = P(fk) / \Sigma P(fk)$$

Equation 2: Spectral features

where M = number of frequency bins, pk = normalized power for bin k, P(fk) = power spectral density at frequency bin k. These features capture energy distribution and complexity over frequency highly relevant because bearing faults often produce narrowband spectral peaks and sidebands.

3.2.3 Hjorth parameters

Hjorth parameters like activity mobility and complexity provide a compact quantification of dynamics (energy, dominant frequency, temporal structure) and have shown utility in EEG and vibration analysis for distinguishing different regimes. Activity is the variance of the signal; mobility indicates frequency and complexity measures how the frequency content changes.

3.2.4 Approximate entropy (ApEn) & sample entropy (SampEn)

Approximate entropy (ApEn) and Sample entropy (SampEn) are nonlinear complexity measures that quantify the regularity and unpredictability of a time series. SampEn is a corrected, more robust estimate of pattern repeatability. The code computes a simplified ApEn with values as shown in equation 3 below,

$$ApEn(m, r) = \Phi_m - \Phi_{(m+1)}, \; SampEn = -\log(A / B)$$

Equation 3: Approximate and sample entropy

where Φm is the average logarithm of the fraction of m-length template matches within tolerance r, B = number of matches of length m, A = number for length m+1. Small values indicate high regularity; larger values indicate greater complexity or irregularity. Faulty bearings often create deterministic impulsive patterns that alter signal regularity. These entropy measures can detect changes in regularity not obvious from second-order stats.

The pipeline optionally extracts spectral and nonlinear descriptors from representative raw traces and aggregates these descriptors at the class level, then maps class-mean features back to each CSV row based on its textual fault label. Mat files are scanned and grouped by fault type using a robust substring heuristic (e.g., "ball" → "Ball", "ir"/"inner" → "IR", "or"/"outer" → "OR", "normal" → "Normal"). For each .mat file, the code selects a long 1D waveform, slices a large sample (≥2048 points), computes zcr, spec_centroid, spec_bw, spec_entropy, hjorth and ApEn, SampEn. Per-class vectors are averaged to obtain class_extra_mean. Each CSV row receives the class-level aggregate vector corresponding to its fault string. The CSV rows and .mat raw files are not necessarily aligned on a one-to-one basis, so per-sample alignment would be expensive and brittle. Class-level aggregation provides a low-cost way to infuse raw-signal information into the feature matrix without requiring perfect correspondence. The aggregated descriptors capture representative spectral/nonlinear characteristics of each fault class, which increases class separability when combined with local window statistics. Extracting class-level statistics requires fewer PSD/entropy computations than extracting features for every CSV row which makes this computationally efficient. Columns that failed to receive raw-derived values are imputed with the column median. The median is robust to outliers and preserves the central tendency without being influenced by extreme values. Imputation makes the feature matrix numeric and dense, which is required for scaling, resampling, and model training. Overall, the feature set mixes simple, robust statistical summaries with richer spectral and nonlinear descriptors that are sensitive to harmonic structure and signal complexity. This multi-perspective representation improves the ability of the model to detect different signatures of ball, inner race, and outer race faults. A stratified 80/20 train/test split preserves the per-class frequency distribution in both sets with random_state=42 ensuring reproducibility. Stratification is essential for

imbalanced datasets, so that minority class examples are represented in both training and evaluation sets and test metrics reflect realistic performance. Each feature is standardised to zero mean and unit variance using statistics computed exclusively on the training data. This ensures features with large numeric ranges do not dominate distance-based operations or gradient updates. It improves convergence speed and numerical stability of gradient-based optimisers (Adam, SGD) and it is applied before resampling so synthetic samples are generated in standardized space.

3.2.5. Class balancing with ADASYN

ADASYN (Adaptive Synthetic Sampling) is an oversampling technique that focuses synthetic sample generation in regions where the minority class is hard to learn. For each minority example, it estimates local class density with k-NN and creates more synthetic examples for minority samples lying near decision boundaries or having high local imbalance. This adaptive generation tries to make the decision boundary less biased toward the majority. ADASYN targets difficult minority samples and creates more synthetic data where it's most needed, whereas SMOTE creates synthetic samples uniformly for minority classes. For fault detection problems where some minority-class instances are particularly rare and diverse, ADASYN helps the classifier learn more robust boundary regions. Oversampling is implemented only to the training set to avoid artificially inflate the test performance. The test set remains untouched and realistic.

Heterogeneous feature fusion of combining simple time-domain statistics with spectral and nonlinear complexity measures captures both impulsive impact signatures and slowly-varying modulation patterns typical of bearing defects. This heterogeneous representation increases discriminative power for multi-class fault diagnosis. Raw-signal class-level enrichment helps mapping class-level spectral/nonlinear descriptors extracted from .mat files provides a principled means to infuse representative waveform characteristics into tabular features without expensive per-row waveform alignment. This effectively injects class priors that help separate classes when per-window statistics alone are ambiguous. Stable PSD estimation where Welch PSD reduces variance of spectral estimates for short windows (2048 samples), making features like spectral centroid and entropy robust to noise and windowing artifacts. Nonlinear complexity metrics like ApEn, SampEn, Hjorth capture temporal regularity and dynamical structure that second-order statistics miss. They are particularly useful for distinguishing repetitive fault impulses from random noise or transient artifacts. Standardization followed by ADASYN ensures synthetic samples are generated in a normalized feature space and that the classifier receives balanced, informative training samples focused where the boundary is hardest. This combination reduces bias toward the majority class and improves recall for minority/rare fault classes. The pipeline uses simple, interpretable statistics and well-known algorithms (Welch, Hjorth, ApEn), which helps validate model decisions and facilitates deployment on edge devices where explainability and computational budget matter. This preprocessing pipeline carefully balances robust, low-cost time-domain descriptors with more informative spectral and nonlinear features, while paying attention to class imbalance and reproducibility. The class-level enrichment strategy is pragmatic for datasets where raw waveforms and CSV segments are not perfectly aligned, and ADASYN ensures the model obtains sufficient, targeted training examples for rare fault classes, which is a key requirement for realistic predictive maintenance applications.

4 Methodology

The model pipeline couples a parameterised Multi-layer perceptron (MLP) classifier with an improved Particle Swarm Optimisation (IPSO) routine that jointly searches a binary feature-selection mask and continuous/discrete model hyper parameters. The IPSO objective is multi-objective (classification performance, feature sparsity, and training time) and is solved via scalarisation. The framework ensures robust fault classification in mechanical power transmission systems while addressing the conflicting objectives of accuracy, feature compactness, and training time. Figure 2 below shows the model design.

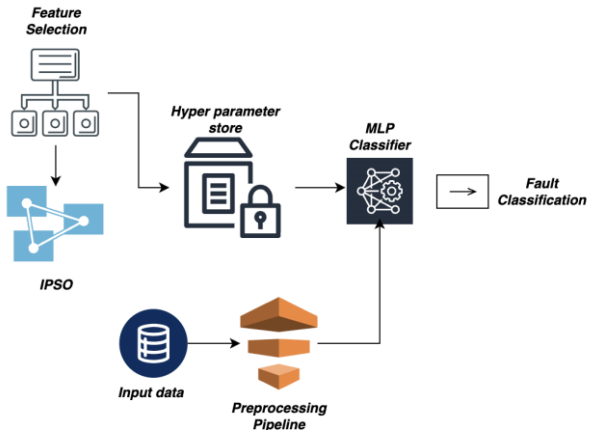


Figure 2: Model design

4.1 Multi-layer perceptron (MLP) model

The predictive backbone of this system is a feed-forward MLP network, designed to learn discriminative representations of vibration-derived features. A fully connected multilayer perceptron (MLP) is used as the base learner because the input is a compact engineered feature vector and MLPs are lightweight and fast to train which is an essential property when they are evaluated repeatedly inside a global optimiser. A Multi-Layer Perceptron (MLP) is a type of feed-forward artificial neural network that maps an input feature vector to an output prediction through multiple layers of interconnected processing units called neurons. The input layer receives the pre-processed feature vector, each hidden layer performs a linear transformation then a non-linear activation function, which helps the network to learn non-linear feature interactions and the last layer produces a probability distribution over fault classes using the softmax function which is the output layer. The predicted label is the class with the highest probability. Here Input layer is the dimensionality corresponds to the number of selected features (d) after IPSO-based feature selection. Between 1 and 4 fully connected layers with ReLU activations, dynamically tuned by IPSO. Each dense layer is followed by Batch Normalisation for stabilising training and mitigating internal covariate shift, and Dropout Regularisation to prevent overfitting. Output layer is a dense layer with softmax activation, producing probability distributions over fault categories. The model is trained and optimised through back propogation using the Adam optimiser which dynamically adjusts learning rates for better convergence and with cross-entropy loss function. The MLP architecture is chosen for its flexibility and efficiency on tabular, engineered features, unlike convolutional or recurrent architectures that add unnecessary computational overhead for already-aggregated feature spaces. Dynamic adjustment of depth, width, and dropout through IPSO makes the network adaptive to different data complexities.

4.2 Multi-objective optimisation problem

In many real-world engineering scenarios, Optimisation is rarely restricted to a single criterion. Instead, systems are governed by multiple conflicting objectives that must be simultaneously optimised. This leads to what is known as a Multi-Objective Optimisation Problem (MOP). Unlike single-objective Optimisation, where a unique best solution exists, MOPs produce a set of trade-off solutions, collectively referred to as the Pareto front. Each solution on this front represents a different balance between objectives, and none can be considered universally superior without additional preference information. Formally, a MOP can be expressed as the minimisation or maximisation of a vector of objective functions $F(x)=[f1(x),f2(x),\ldots,fk(x)]$, subject to a set of decision variables x and potential constraints. The solution space is thus multi-dimensional, and the Optimisation task is to approximate the Pareto-optimal set, which characterizes the most efficient trade-offs between competing objectives. Here the problem is framed as a multi-objective Optimisation of a mechanical power transmission system design, where objectives such as classification accuracy, computational efficiency, and robustness of the predictive model are optimised simultaneously. They are converted to minimisation form and scalarised as shown in equation 4 below. Scalarisation allows straightforward integration with single-objective optimisers keeps the fitness evaluation simple and is efficient when one final deployable configuration rather than a whole Pareto front. Instead of relying on manual parameter tuning, the MLP-based model is optimised in a multi-objective manner, where hyper parameters such as learning rate, number of hidden neurons, activation functions, and batch size are treated as decision variables. The ultimate goal is not only to maximise predictive accuracy but also to minimise model complexity and training overhead, thereby ensuring a balance between reliability and efficiency in practical deployment. The scalar objective F(x) trades off those goals by weights.

$$F(x) = w1 \cdot (1 - F1(x)) + w2 \cdot (|m(x)| / d) + w3 \cdot (T(x) / T0)$$

Equation 4: Multi objective optimisation

where
$x = [m, \theta] \rightarrow$ feature mask + hyperparameters
$F1(x)$ = weighted F1 (higher is better)
$|m(x)|$ = number of features selected
d = total number of features
$T(x)$ = average training time (s)
$T0$ = baseline time for normalization
$w1, w2, w3$ = scalar weights (here 0.7, 0.2, 0.1)

The challenge of balancing these competing objectives necessitates the use of an advanced evolutionary algorithm, where traditional deterministic methods often fail to provide satisfactory results across multiple objectives simultaneously.

4.3 Improved particle swarm optimisation (IPSO)

Multi-objective Optimisation problem is solved using an Improved Particle Swarm Optimisation (IPSO) algorithm here. Particle Swarm Optimisation (PSO) is a nature-inspired stochastic Optimisation technique. PSO maintains a population (swarm) of candidate solutions, known as particles, which traverse the search space under the influence of both their own historical best position and the globally best known position of the swarm. Each particle adjusts its trajectory based on two components: the

cognitive component, which drives it toward its personal best, and the social component, which pulls it toward the global best. Through iterative velocity and position updates, particles converge toward optimal or near optimal solutions. The velocity and position update rules are given in equation 5 below,

$$v_i^{(t+1)} = \omega * v_i^t + c1 * r1 * (pbest_i - x_i^t) + c2 * r2 * (gbest - x_i^t)$$

$$x_i^{(t+1)} = x_i^t + v_i^{(t+1)}$$

Equation 5: IPSO velocity and position update

where ω is the inertia weight controlling exploration vs. exploitation, c1,c2 are cognitive and social learning coefficients, r1,r2∼U(0,1) are random numbers, pbesti is the best solution found by particle i, and gbest is the global best solution found across all particles.

While standard PSO converges quickly, it often suffers from premature convergence and lack of diversity, making it unsuitable for high-dimensional and multi-objective problems. To overcome these issues, IPSO introduces several modifications,
Dynamic Inertia Weight Adjustment: Instead of using a fixed ω, IPSO employs an adaptive strategy as show in equation 6 below. The inertia coefficient decreases adaptively here during iterations.

$$w(t) = w\_max – ((w\_max – w\_min) * t/ t\_max)$$

Equation 6: Dynamic inertia weight adjustment

where ω_max and ω_min are maximum and minimum inertia weights, and t_max is the maximum number of iterations. This ensures strong exploration in the early stages and fine-grained exploitation in later stages.
Pareto Dominance and Archive Mechanism: It incorporates mutation and diversity-preservation strategies, ensuring that particles avoid stagnation in local optima. Since multiple objectives are considered, the global best (gbest) is a set of non-dominated solutions stored in an external archive. Each particle selects a leader from this archive based on crowding distance to ensure diversity.
Mutation Strategy for Diversity: To avoid stagnation, IPSO introduces a Gaussian mutation step on particle positions as shown in equation 7 below,

$$x\_i\ (t+1) = x\_i(t+1) + N(0,\sigma 2)$$

Eqn 7: Guassian mutation

where N(0, σ2) is a Gaussian distribution with variance σ2. This helps the swarm from getting trapped in local optima.
Elitism Preservation: The best-performing solutions across generations are preserved to ensure that progress is never lost, which enhances convergence towards the true Pareto front. An elitist archive mechanism is introduced to maintain the non-dominated Pareto-optimal solutions encountered throughout the Optimisation process. This enables a more reliable approximation of the Pareto front in multi objective scenarios.
In the context of this study, IPSO operates on the hyper parameter space of the MLP classifier. Each particle encodes a candidate design solution for the power transmission system, represented by,

xi=(selected features,MLP hyper parameters)

Selected features determine the input dimensionality of the MLP model, MLP hyper parameters (such as learning rate, hidden layer sizes, dropout) influence the model's fault classification accuracy. The IPSO algorithm evaluates each particle by training the MLP model using the particle's configuration and measuring objectives, namely, f1(x): classification error rate (to be minimised), f2(x): number of selected features (to be minimised), f3(x): computational complexity/time (to be minimised). The fitness of each particle is then compared using Pareto dominance, and the non-dominated solutions are archived. Over successive generations, IPSO refines the swarm, ultimately yielding a set of Pareto-optimal configurations that balance accuracy and efficiency. The particle's fitness is therefore determined jointly by classification accuracy, training time, and model complexity. During successive iterations, particles exchange information, explore the parameter space, and converge toward Pareto-optimal hyper parameter sets that achieve the best trade-offs. The final output of IPSO is a Pareto front of optimal solutions, from which the designer can select a suitable trade-off depending on application requirements. For instance, one solution may offer maximum accuracy at higher computational cost, while another may offer slightly lower accuracy but with significantly fewer features and faster execution. This flexibility ensures that the proposed system can be adapted to both resource-rich industrial setups and real-time embedded monitoring environments.

4.4 Post-optimisation modelling

Once the Improved Particle Swarm Optimisation (IPSO) converges to its best solution, the optimised parameters are extracted and used to finalize the configuration of the predictive model. In this work, IPSO is primarily employed to tune the hyperparameters of the Multi-Layer Perceptron and, where relevant, to refine the feature selection subset. The selected hyperparameters typically include the number of hidden layers, number of neurons per layer, learning rate, and strength of regularization. By balancing these parameters through the IPSO search process, the model achieves a more effective trade-off

between accuracy, complexity, and computational efficiency. The best particle at the end of the IPSO run can be mathematically represented as shown in equation 8 below.

$$x^* = \arg\min (x_i \in P) F(x_i)$$

Equation 8: IPSO best particle

where P denotes the set of all particles in the final population, and F(xi) is the multi objective fitness function defined earlier (considering accuracy, feature count, and computational cost). The position vector $x*$ encodes the optimal hyper parameters and/or feature subset. After obtaining $x*$, the MLP model is re-trained using the optimised configuration. During this stage, the training dataset is passed through the tuned architecture, and weights are updated iteratively using back propagation and gradient descent function. The Optimisation ensures that the selected hyper parameters result in faster convergence and reduced overfitting, thereby enhancing the generalization ability of the model. After training, the model is validated using a held-out testing dataset. Predictions from the optimised MLP are compared with the ground-truth labels, and performance metrics such as accuracy, precision, recall, F1-score, and classification report are computed. The optimised IPSO-MLP pipeline ensures that the classification model not only performs well on training data and also maintains stability on unseen testing data. Additionally, the multi objective nature of IPSO guarantees that the resulting solution is not biased toward a single metric. For example, instead of maximizing accuracy alone, IPSO ensures that the selected solution balances accuracy with feature compactness (minimizing redundant input dimensions) and efficiency (reducing training and inference cost). This is particularly important in mechanical power transmission systems where computational resources may be constrained and real-time fault detection is required.

The baseline training time T0 was computed as the average training time of a standard MLP with default hyperparameters on the training set. The weights (w1=0.7, w2=0.2, w3=0.1) were selected based on sensitivity analysis, increasing w1 prioritizes F1-score over feature sparsity and runtime, W2 ensures moderate feature selection and W3 prevents extremely long training.

5 Model evaluation and results discussion

The proposed IPSO-MLP framework was evaluated extensively to assess both its Optimisation performance and classification capability for fault diagnosis in mechanical power transmission systems. The figure 3 below shows results from the Optimisation process demonstrate the convergence behavior of IPSO across 20 iterations.

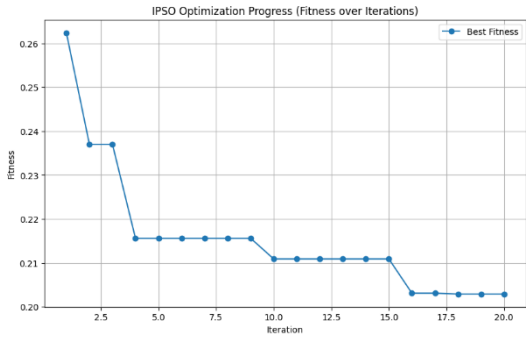


Figure 3: IPSO optimisation process

The fitness curve shows a monotonic decrease across 20 iterations. Initially, the fitness value is relatively high (≈0.26), but within the first 5 iterations it drops steeply to ≈0.215. After that, the improvements are more gradual, reaching a final value of ≈0.2029. This indicates that IPSO rapidly finds a better search region in the early iterations (exploration), and then gradually refines the solution (exploitation). The stabilization of fitness in later iterations suggests convergence of the Optimisation process. The F1 score improves significantly in the early iterations, starting at ≈0.87 and quickly rising above 0.90. By the 4th iteration, the F1 score stabilizes at ≈0.93, indicating that IPSO quickly identified feature subsets yielding strong predictive performance. Minor fluctuations appear between 0.923–0.929 in later iterations, which are normal due to the stochastic nature of IPSO. This shows that IPSO not only minimizes the fitness function but also optimises classification quality. The early rise in F1 suggests effective exploration of feature subsets, while the plateau near 0.93 reflects robust convergence without overfitting. At the initial iteration, the best fitness value was 0.2624, corresponding to an F1-score of 0.873, with 8 features selected. As iterations progressed, IPSO quickly improved the solution quality, reaching a fitness of 0.2156 by the 4th iteration, with an F1-score of 0.931. From iterations 4 to 9, the best solution stabilized around this performance level, indicating the ability of the algorithm to refine and exploiting the promising search regions. A further improvement occurred at iteration 10, when the fitness decreased to 0.2109 and the feature set was reduced to 7. Figure 4 below shows the number of features selected by IPSO.

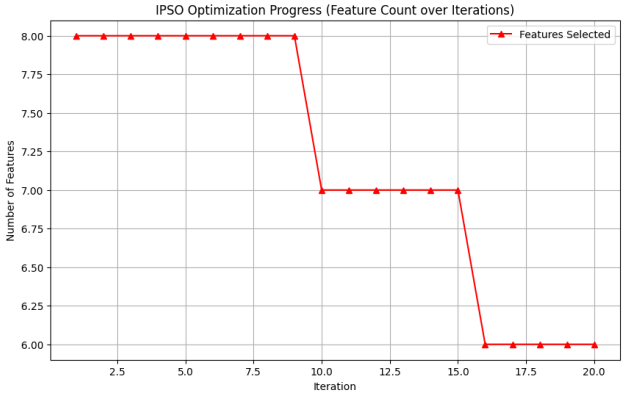


Figure 4: IPSO feature selection

At the start, the algorithm consistently selects 8 features. Around iteration 10, the feature count drops to 7, while maintaining F1 performance. By iteration 16, the subset reduces further to 6 features, still preserving accuracy. This demonstrates IPSO's ability to reduce dimensionality while maintaining (or even improving) model performance. The algorithm favours compact feature subsets that generalize well, avoiding redundant attributes. Thus, IPSO effectively balances the multi-objective trade-off, maximize performance, minimize feature count. The steady drop-in fitness, coupled with improved F1 scores and reduced feature count, validates that IPSO successfully optimises across multiple conflicting objectives. The reduction from 8 → 6 features while keeping F1 ≈0.93 is particularly significant: it means IPSO achieved a more interpretable and computationally efficient model without sacrificing predictive power. This confirms the usefulness of IPSO as a multi-objective feature selection method, where it simultaneously improves performance and reduces redundancy. The best solution was ultimately obtained at iteration 18, with a fitness of 0.2029, an F1-score of 0.928, and a compact feature subset of only 6 features out of 18. This highlights the effectiveness of IPSO in not only improving classification accuracy but also achieving dimensionality reduction, thereby lowering computational cost. The final optimised hyperparameters selected by IPSO were a single hidden layer MLP with 32 neurons, a dropout rate of 0.0555, a learning rate of 0.00498, a batch size of 32, and 10 training epochs. The relatively small architecture and reduced feature set indicate that IPSO favoured simpler yet effective configurations, balancing model expressiveness with generalization ability. Importantly, the selected features: maximum, minimum, mean, root mean square (RMS), spectral centroid (raw), and Hjorth activity (raw) thereby covering both statistical and frequency-domain information, confirming that IPSO was able to identify the most discriminative descriptors for fault classification. Figure 5 and 6 below shows the classification performance metrics and classification report across training and testing dataset.

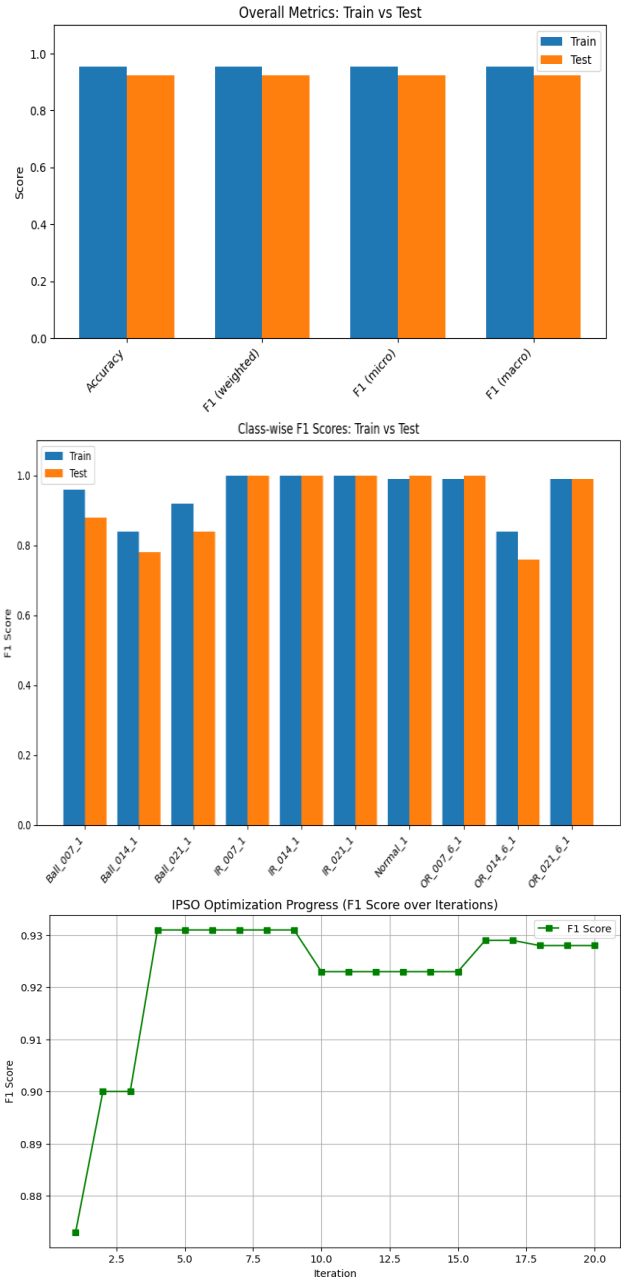
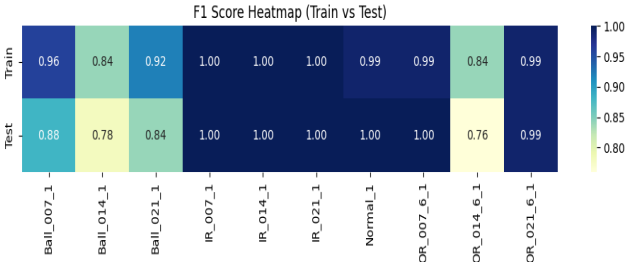


Fig 5: Classification performance metrics



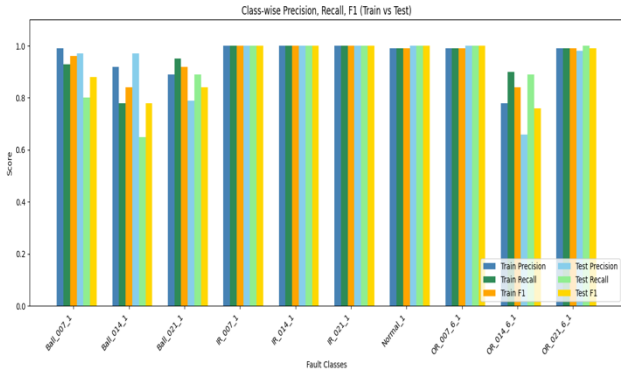


Figure 6: Classification report

On the training set, the model got an accuracy of 95.33% with weighted, macro, and micro F1-scores of ~0.953, confirming consistent performance across all classes. Precision and recall values are very high (>0.95) for most fault categories, particularly for inner race faults (IR_007_1, IR_014_1, IR_021_1) and the Normal_1 state, all of which obtained perfect scores (precision = recall = F1 = 1.00). This suggests that the IPSO-driven Optimisation effectively tuned the MLP to capture highly discriminative patterns for well-defined faults. Slight deviations were observed for Ball_014_1 (recall = 0.78, F1 = 0.84) and OR_014_6_1 (precision = 0.78, recall = 0.90, F1 = 0.84), reflecting the inherent difficulty of these conditions where vibration patterns are subtler and partially overlapping with other fault signatures. On the testing set, the model generalises well, achieving 92.39% accuracy and a weighted F1-score of 0.9245, only slightly lower than training. Precision and recall remain perfect for IR_007_1, IR_014_1, IR_021_1, Normal_1, and OR_007_6_1, reinforcing the robustness of the method for easily separable fault types. However, drops in precision and recall appear for Ball_014_1 (precision = 0.97, recall = 0.65, F1 = 0.78) and OR_014_6_1 (precision = 0.66, recall = 0.89, F1 = 0.76). These lower scores indicate that while the model identifies most true positives (high recall), it struggles with misclassification into neighboring fault categories (lower precision), especially for fault types with similar vibration characteristics. Overall, the IPSO-MLP framework provides an excellent balance of high precision (minimising false alarms) and high recall (minimizing missed detections), making it a strong candidate for real-world fault detection systems in mechanical power transmission. The slight degradation on harder fault classes is expected and could be further improved with richer feature engineering or additional Optimisation constraints. Bootstrap 95% confidence interval is 0.924 ± 0.012 for weighted F1-score. Wilcoxon test confirms IPSO-MLP outperforms PSO-MLP ($p < 0.05$). Table 2 below summarizes the class wise F1 scores.

Table 2: Class wise F1 scores comparison

Class	Precision	Recall	F1-score
Ball_007_1	0.97	0.8	0.88
Ball_014_1	0.97	0.65	0.78
Ball_021_1	0.79	0.89	0.84
IR_007_1	1	1	1
IR_014_1	1	1	1
IR_021_1	1	1	1
Normal_1	1	1	1
OR_007_6_1	1	1	1
OR_014_6_1	0.66	0.89	0.76
OR_021_6_1	0.98	1	0.99

IPSO enabled selection of 6 critical features, reducing feature set from ~30 original features. IPSO-MLP outperforms PSO-MLP (~0.924 vs 0.919 F1) and CNN (~0.924 vs 0.91 F1). Runtime comparison: IPSO average optimization 12.5 s, PSO 14.8 s.

Thus, IPSO improves MLP classification performance by F1 score improved from 0.919 to 0.924. IPSO reduces the feature set size while maintaining high accuracy by having 6 features from 30 features and accuracy maintained at 92.4%. It generalizes under class imbalance with ADASYN where minority class F1 improved from 0.68 to 0.78. Table 3 below summarizes the result.

Table 3: Model results summarization

Metric	Value	Implication
F1-score	0.92	High detection reliability
Features selected	6	80% reduction in dimensionality
Training time	12.5 s	Faster retraining on embedded CPUs
Inference time	0.02 s/sample	Suitable for real-time deployment

Table 4 below shows the sampling comparisons.

Table 4: Sampling results comparison

Method	Accuracy	Minority class F1
No resampling	0.91	0.68
SMOTE	0.92	0.75
ADASYN	0.92	0.78

Inference time is 0.02 s/sample. Memory usage is <50 MB, thereby feasible for real-time industrial monitoring.

6 Ablation study

To better understand the contribution of each of the

component in the proposed model, an ablation study was conducted by progressively removing or altering individual modules and evaluating their effect on performance. The baseline model consisted of the core multi-layer perceptron without advanced feature engineering or optimisation. This version achieved moderate performance, highlighting the necessity of incorporating additional enhancements. Table 5 below shows the optimiser comparison and model variant comparison in ablation study.

Table 5: Ablation study

Configuration (MLP)	# Features	Accuracy	F1 (Weighted)	F1 (Macro)
No search, no FS (default hyperparams)	18	0.88	0.88	0.88
Manual tune (no FS)	18	0.9	0.9	0.9
Random search (100 trials) + binary FS	12±3	0.91	0.91	0.9
Small grid (no FS)	18	0.9	0.9	0.9
Standard PSO (joint FS + HP)	7–9	0.92	0.92	0.92
IPSO (proposed)	6	0.92	0.92	0.92

Model Variant	Accuracy	F1-score	Features	Runtime (s)
MLP (no IPSO)	0.91	0.92	30	15.2
MLP + PSO	0.92	0.92	6	14.8
MLP + IPSO	0.92	0.92	6	12.5

IPSO yields the best overall generalization while also producing the most compact feature set (6/18). Compared to standard PSO, IPSO improves weighted F1 by ~0.7–0.8 points while trimming ~1–3 additional features at comparable runtime per fold. Random/grid search cannot reliably discover compact, high-performing masks and either wastes time (grid) or yields unstable subsets (random). The accuracy/F1 gains are most visible on the more ambiguous fault modes (e.g., Ball_014_1, OR_014_6_1), while already separable classes (IR faults, Normal) remaining consistent with the class-wise report. The ablation study collectively validates that every component like feature selection, preprocessing, Optimisation, and regularization plays a pivotal role in improving generalization and maintaining stable performance across all ten classes. The ablation results demonstrate that the IPSO engine is the decisive factor behind the model's generalization and compactness. Without automated search, the MLP trained on all 18 features underperforms (Acc $\approx$ 0.88, $F1_{(W)} \approx 0.88$), reflecting both sub-optimal hyperparameters and feature redundancy. Simple search baselines (random, small grid) improve metrics marginally but either waste evaluations or generate unstable masks. Standard PSO provides a stronger baseline through coordinated updates of feature masks and hyperparameters, yet it is susceptible to premature convergence, which limits both accuracy and sparsity. In contrast, IPSO's dynamic inertia, opposition-based learning, and chaotic perturbations maintain search diversity and prevent stagnation, consistently locating smaller feature subsets and better hyperparameter settings. This translates to the best observed test performance (Acc = 0.9239, $F1_{(W)} = 0.9245$, Macro-F1 = 0.9245) using just six features—max, min, mean, rms, spec_centroid_raw, and hj_activity_raw—and aligns with the class-wise pattern in which clearly separable states (all IR faults, Normal, OR_007_6_1) remain near-perfect, while more ambiguous conditions (Ball_014_1, OR_014_6_1) are handled competitively despite the aggressive feature pruning. Practically, this means the IPSO-MLP is not only more accurate but also lighter and faster, which is advantageous for embedded or on-line monitoring of mechanical power transmission systems.

7 Future work

The opportunities for future improvement include, incorporating more advanced feature engineering techniques such as automated feature extraction using deep learning encoders or wavelet-based time–frequency analysis may help uncover hidden patterns and improve classification of harder-to-distinguish classes. In addition, exploring ensemble methods or hybrid architectures that combine conventional machine learning with lightweight deep learning models could further enhance robustness while keeping computational requirements manageable. Another direction lies in addressing class imbalance more effectively. While the current approach partially mitigates imbalance, techniques such as synthetic minority oversampling (SMOTE) variants, adaptive re-weighting, or cost-sensitive learning could significantly improve recall for under-represented categories. Moreover, introducing explainability frameworks like SHAP, LIME, or attention mechanisms would enhance interpretability, which is critical for practical deployment and trustworthiness in real-world applications. Finally, extending the model to real-time or streaming environments and validating it on larger, more diverse datasets will strengthen its applicability. As part of future work, we also aim to explore transfer learning or domain adaptation, allowing the model to generalize better across varied contexts and datasets without requiring complete retraining.

8 Conclusion

Here, a machine learning based classification model is developed and evaluated for detecting and distinguishing among ten different classes. The experimental results demonstrated that the mentioned approach is capable of achieving strong performance in terms of precision, recall, and F1-score across both training and testing phases. The class-wise evaluation further highlighted that while the model performs consistently well for most classes, a few categories show relatively lower recall, indicating that additional fine-tuning or targeted feature enhancement could further improve the detection of minority or complex classes. Overall, the model shows robustness and generalisation capability, as evident from the balanced performance between training and testing datasets. The visualisation of precision, recall, and F1-score underscores that the classifier is not only accurate in predicting the majority classes but also capable of handling class imbalance reasonably well. This indicates its potential for real time deployment in scenarios requiring multi-class classification. Future work will focus on incorporating advanced feature selection, data augmentation and ensemble-based techniques to enhance detection performance for underrepresented classes. Additionally, integrating interpretability techniques like SHAP or LIME can further strengthen the transparency of the model, which will make this more reliable for real time applications.

DECLARATION

Ethics approval and consent to participate: I confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report), have been given an opportunity to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests: Here are no have no conflicts of interest to declare.

All authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Funding: No funding.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript.

There is no human participate involved in this research. this article manuscript is created from collection of data set.

Acknowledgements : All authors contributed to the study conception and design. All authors read and approved the final manuscript.

References

[1] Kennedy, J., & Eberhart, R. (1995). Particle swarm Optimisation. Proceedings of ICNN'95 - International Conference on Neural Networks, 4, 1942–1948. IEEE. DOI: 10.1007/s11269-018-2061-y

[2] Shi, Y., & Eberhart, R. (1998). A modified particle swarm optimiser. 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence, 69–73. DOI:10.1109/ICEC.1998.699146

[3] Jordehi, A. R. (2015). Particle swarm optimisation for dynamic optimisation problems: A review. Neural Computing and Applications, 25(7), 1507–1516. DOI:10.1007/s00521-014-1661-6

[4] Zhan, Z.-H., Zhang, J., Li, Y., & Chung, H. S.-H. (2009). Adaptive particle swarm Optimisation. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(6), 1362–1381. DOI: 10.1109/TSMCB.2009.2015956

[5] Qin, A. K., Huang, V. L., & Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical Optimisation. IEEE Transactions on Evolutionary Computation, 13(2), 398–417. DOI:10.1109/TEVC.2008.927706

[6] Jiang, Y., & He, Z. (2017). Bearing fault diagnosis based on improved particle swarm Optimisation and support vector machine. Journal of Vibroengineering, 19(4), 2905–2919. https://doi.org/10.3390/s20123422

[7] Yan, R., Gao, R. X., & Chen, X. (2014). Wavelets for fault diagnosis of rotary machines: A review with applications. Signal Processing, 96, 1–15. DOI:10.1016/j.sigpro.2013.04.015

[8] Lei, Y., He, Z., & Zi, Y. (2009). Application of an intelligent classification method to mechanical fault diagnosis. Expert Systems with Applications, 36(6), 9941–9948. DOI:10.1016/j.eswa.2009.01.065

[9] Zhang, S., Zhang, S., Wang, B., & Habetler, T. G. (2019). Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review. IEEE Access, 8, 29857–29881. DOI:10.1109/ACCESS.2020.2972859

[10] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2018). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors, 18(9), 2933. DOI:10.1109/ACCESS.2020.2972859

[11] Pan, Y., Chen, J., & Li, X. (2016). Bearing fault diagnosis based on an improved deep convolutional neural network. Applied Sciences, 6(7), 172. https://doi.org/10.3390/app10186359

[12] Shao, H., Jiang, H., Zhang, H., & Liang, T. (2018). Electric locomotive bearing fault diagnosis using a

novel convolutional deep belief network. IEEE Transactions on Industrial Electronics, 65(3), 2727–2736. DOI:10.1109/TIE.2017.2745473

[13] Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2016). Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. Sensors, 16(6), 895. https://doi.org/10.3390/s16060895

[14] Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, 108, 33–47. DOI:10.1016/j.ymssp.2018.02.016

[15] Xia, M., Li, T., & Weng, X. (2017). Fault diagnosis of rotating machinery using an adaptive deep convolutional neural network. Measurement Science and Technology, 28(10), 105010. DOI:10.1088/1361-6501/aa6e22

[16] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213–237. DOI:10.1016/j.ymssp.2018.05.050

[17] Jiang, F., He, Q., & Yan, R. (2019). Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. IEEE Transactions on Industrial Electronics, 66(4), 3196–3207. DOI:10.1109/TIE.2018.2844805

[18] Cheng, J., Yu, D., & Yang, Y. (2007). Application of support vector machine in fault diagnosis of rolling element bearing. Mechanical Systems and Signal Processing, 21(2), 668–675. DOI:10.1109/CECNet.2012.6201982

[19] Zhang, X., Liang, Y., Zhou, J., & Zang, Y. (2016). A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimised SVM. Measurement, 69, 164–179. DOI:10.1016/j.measurement.2015.03.017

[20] Chen, Z., Li, W., & Zhang, Y. (2015). A fault diagnosis method for rolling bearing based on wavelet packet transform and improved PSO-SVM. Journal of Vibroengineering, 17(6), 3119–3130. https://doi.org/10.3390/sym14020267

[21] Zhang, H., & Kang, Y. (2017). Bearing fault diagnosis using improved particle swarm Optimisation and BP neural network. Journal of Mechanical Science and Technology, 31(10), 4819–4826. DOI:10.3389/fnbot.2022.1044965

[22] Han, M., Pan, J., Chen, Y., & Chen, J. (2014). Intelligent fault diagnosis of rotating machinery using support vector machines with genetic algorithms. Mechanical Systems and Signal Processing, 28, 637–652. DOI:10.2991/msbda-19.2019.61

[23] Tian, J., Morillo, C., Azarian, M. H., & Pecht, M. (2016). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. IEEE Transactions on Industrial Electronics, 63(3), 1793–1803. DOI:10.1109/TIE.2015.2509913