Noise-Robust Bearing Fault Diagnosis via Enhanced EFD and QPSOL-Optimized SVM.

HungLinh Ao [1]*, BaoQuoc Doan[1], HoaiQuoc Le [1]

[1] Faculty of Mechanical Engineering, Industrial University of HCM City, HCM City, Viet Nam
*e-mail: aohunglinh@iuh.edu.vn

Abstract— This paper presents an optimization-driven framework for noise-robust bearing fault diagnostics aimed at enhancing the reliability and design performance of rotating machinery systems. The proposed approach integrates an enhanced empirical ensemble Fourier decomposition (EEFD) with a support vector machine (SVM) whose hyperparameters are optimized using a quadratic interpolation particle swarm optimization with local search (QPSOL) algorithm.
To address signal degradation under harsh industrial environments, EEFD is employed to decompose vibration signals and extract high-quality intrinsic components. A compact yet discriminative multi-domain feature set, including Root Mean Square (RMS), Kurtosis, and Hjorth Mobility (HM), is constructed to characterize the dynamic behaviour of the system. The QPSOL algorithm is then utilized to optimize the SVM parameters, improving convergence accuracy and classification robustness.
Experimental validation under a 10 dB signal-to-noise ratio demonstrates that the proposed method achieves superior diagnostic accuracy and convergence performance compared with conventional optimization techniques. Beyond fault classification, the developed framework provides a basis for simulation-driven design optimization and reliability-oriented decision-making in mechanical systems, enabling more effective predictive maintenance strategies and lifecycle performance improvement.

Keywords: Optimization, Simulation-driven design, Bearing fault diagnosis, QPSOL algorithm, Support Vector Machine (SVM), Empirical Ensemble Fourier Decomposition (EEFD), Reliability engineering.

I. INTRODUCTION

Rolling element bearings are essential components in rotating machinery, serving as the primary interface enabling smooth relative motion between rotating shafts and stationary housings. Their operational condition directly influences the reliability and performance of the entire mechanical system. Industrial studies indicate that bearing-related failures account for approximately 40–50% of total failures in rotating equipment [1]. If not detected at an early stage, such failures may result in severe system degradation, substantial economic losses, and safety risks.

When localized defects, such as cracks, spalling, or surface wear, occur on the inner race (IR), outer race (OR), or rolling elements, they disrupt the regular motion and generate characteristic vibration signatures. These defects are typically associated with specific fault frequencies, including the Ball Pass Frequency of the Outer race (BPFO), Ball Pass Frequency of the Inner race (BPFI), and the Ball Spin Frequency (BSF). While these features are clearly distinguishable under controlled laboratory conditions, their identification in practical industrial environments remains challenging due to strong background noise, complex transmission paths, and varying operating conditions. As a result, traditional diagnostic methods often exhibit limited robustness and reduced accuracy in early fault detection.

In modern engineering practice, such limitations extend beyond condition monitoring and directly impact simulation-driven design and system optimization processes. Reliable diagnostic models can be integrated into engineering workflows to support reliability-based design, predictive maintenance strategies, and lifecycle performance optimization. Therefore, the development of robust and noise-resilient diagnostic methodologies represents a critical step toward multidisciplinary design optimization in mechanical systems.

To address these challenges, researchers have explored various signal processing techniques, including Fast Fourier Transform (FFT), Empirical Mode Decomposition (EMD), and Variational Mode Decomposition (VMD). Recently, the combination of Empirical Fourier Decomposition (EFD) and Singular Value Decomposition (SVD) with Support Vector Machines (SVM) has gained attention. For instance, Ao et al. (2020) proposed the EFD-SVD-HGS-SVM method, which achieved an error rate of 5.68% on the CWRU dataset [2]. Despite its success, this approach relies heavily on SVD for dimensionality reduction, which adds computational complexity, and utilizes the Hunger Games Search (HGS) algorithm, which often requires significant optimization time to reach convergence.

Feature selection also plays a crucial role in enhancing diagnostic sensitivity. In addition to classical statistical indicators, Hjorth parameters have proven effective in capturing high-frequency and non-linear signal characteristics [3]. Meanwhile, optimization algorithms have been increasingly employed to improve model performance. Among them, Quadratic Particle Swarm Optimization with Local Search (QPSOL) has demonstrated superior convergence properties by combining global search capability with local refinement strategies [4]. Despite its effectiveness in general optimization problems, its potential for improving engineering system performance and supporting optimization-driven design processes in fault diagnosis applications has not yet been fully explored.

This study addresses these limitations by proposing an integrated optimization-oriented diagnostic framework, denoted as EEFD-RMS-Kurtosis-HM-QPSOL-SVM. The framework incorporates an enhanced Empirical Ensemble Fourier Decomposition (EEFD) technique for robust noise suppression, together with a compact set of discriminative time-domain features (RMS, Kurtosis, and Hjorth Mobility). The QPSOL algorithm is employed as a core component to optimize the SVM hyperparameters, ensuring improved classification performance with reduced computational cost. In addition, the proposed methodology is designed to support simulation-driven engineering workflows, where diagnostic outputs can inform system optimization and reliability-oriented design decisions.

Despite the proven efficiency of QPSOL in solving complex global optimization problems, its application to rolling element bearing fault diagnosis remains largely unexplored. More importantly, its potential to enhance engineering system performance and support optimization-driven design processes in such applications has not yet been systematically investigated.

Conventional SVM-based diagnostic frameworks typically rely on standard metaheuristics such as Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and Hunger Games Search (HGS) for hyperparameter tuning. However, these methods often face limitations when dealing with high-dimensional and noise-corrupted vibration data, leading to suboptimal classification performance.

To the best of our knowledge, the integration of QPSOL with SVM for optimizing the penalty parameter C and kernel parameter γ in bearing fault identification has not been reported in the existing literature. This gap is particularly significant, as the local search capability of QPSOL provides a refined optimization mechanism that can improve parameter accuracy, thereby enhancing diagnostic robustness under industrial noise conditions and supporting more reliable engineering decision-making.

This study addresses the limitations of existing approaches by proposing an integrated, optimization-oriented diagnostic framework, namely EEFD-RMS-Kurtosis-HM-QPSOL-SVM. An enhanced version of Empirical Fourier Decomposition, referred to as Empirical Ensemble Fourier Decomposition (EEFD), is adopted to improve noise suppression and signal quality under adverse operating conditions.

In contrast to conventional approaches that rely on computationally intensive dimensionality reduction techniques such as Singular Value Decomposition (SVD), this work introduces a compact set of highly discriminative time-domain features, including Root Mean Square (RMS), Kurtosis, and Hjorth Mobility (HM), to effectively characterize system dynamics.

Furthermore, the QPSOL algorithm is employed as a core optimization component to automatically tune the hyperparameters of the SVM, improving convergence efficiency and model robustness. The proposed framework is designed not only to enhance diagnostic accuracy but also to support reliability-oriented engineering analysis and optimization-driven decision-making in mechanical systems. The main contributions of this work are summarized as follows:

- Enhanced Noise Robustness: An improved EEFD-based signal decomposition approach is developed to effectively suppress Gaussian noise and enhance signal quality under low signal-to-noise ratio (SNR) conditions.
- Compact and Efficient Feature Representation: A simplified yet highly discriminative feature set (RMS–Kurtosis–HM) is proposed, eliminating the need for computationally expensive dimensionality reduction techniques such as SVD while preserving diagnostic sensitivity.
- Optimization-Driven Model Development: The QPSOL algorithm is employed as a core optimization strategy to tune SVM hyperparameters, achieving faster convergence and improved robustness compared with conventional metaheuristic approaches.
- Superior Performance and Efficiency: The proposed framework demonstrates enhanced classification accuracy and computational efficiency relative to state-of-the-art models, including HGS-SVM, SHO-SVM, and WOA-SVM.

II. Enhanced Empirical Fourier Decomposition – EEFD

EEFD is a recently proposed nonlinear, non-stationary signal decomposition method that combines the ideas of Empirical Wavelet Transform (EWT) and Fourier Decomposition Method (FDM). This method has a complete mathematical foundation, yielding accurate and consistent decomposition results with higher computational efficiency compared to traditional EMD, VMD, or EWT [5].

EEFD operates based on zero-phase filters to partition the frequency spectrum and reconstruct mode functions (MFs) in the time domain. The basic EEFD procedure consists of the following steps:

A. Empirical Fourier Decomposition (EFD)

Empirical Fourier Decomposition (EFD) is an adaptive signal decomposition technique that combines the advantages of Fourier-based analysis and data-driven segmentation. It decomposes a nonstationary signal into a set of mono-components, referred to as mode functions (MFs), each representing a specific frequency band.

Given a vibration signal $x(t)$, its Fourier spectrum $X(\omega)$ is first obtained via the Fourier transform. The frequency range $[0, \pi]$ is then partitioned into multiple subbands based on the local extrema of the amplitude spectrum, the corresponding frequency boundaries are determined as:

$$v_n = \arg\min X(\omega), \Omega_n < \omega < \Omega_{n+1} \tag{1}$$

Based on these boundaries, a set of zero-phase filter banks is constructed as:

$$b_n(\omega) = \begin{cases} 1 & v_{n-1} \le |\omega| \le v_n \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Each subband signal is then obtained by applying the corresponding filter in the frequency domain, followed by the inverse Fourier transform:

$$x_n(t) = \sum_{n=1}^{N} x_n(t) \tag{3}$$

Despite its effectiveness, EFD suffers from several limitations. First, the decomposition result is highly sensitive to noise, especially in practical bearing signals with strong background interference. Second, improper band division may lead to excessively narrow frequency bands, resulting in meaningless components. Finally, the number of decomposition modes is difficult to determine adaptively, which may introduce redundant or irrelevant components.

The mathematical definition is:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-iwt}dt \tag{4}$$

$F(\omega)$ represents the Fourier Transform of the time-domain signal $f(t)$ at angular frequency ω. The transform employs the complex exponential e^{-iwt} to measure how much the signal resonates with a specific frequency component."

The power of this formula is its inverse:

$$f(t) = \frac{1}{2\pi} F(\omega)e^{iwt}d\omega \tag{5}$$

This says that if you take all the frequency components $F(\omega)$ and sum them up (integrate), you get your original signal $f(t)$ back. The transform is lossless

B. Construction of zero – phase filter number

The zero-phase filter was used as bandpass filter in each frequency band without phase change. The Fourier transform of the signal $f(t)$ is given as the following equation:

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t}\,dt \tag{6}$$

The $\mu_n(\omega)$ was used to form a zero phase filter bank by with cut-off frequencies is ω_{n-1} and ω_n :

$$\mu_n(\omega) = \begin{cases} 1 & \text{if } \omega_{n-1} \le |\omega| \le \omega_n \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $1 \le n \le N$ and the values of ω_n are presented by Eq. (1).

The filtered signals corresponding to $\mu_n(\omega)$ are calculated by

$$f_n(\omega) = \mu_n(\omega)f(\omega) = \begin{cases} f(\omega) & \text{if } \omega_{n-1} \le |\omega| \le \omega_n \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The inverse Fourier transform is used to decompose the components in the time domain:

$$f_n(t) = \int_{-\infty}^{+\infty} f_n(\omega)e^{j\omega t}\,\mathrm{d}\omega = \int_{-\omega_n}^{-\omega_{n-1}} f_n(\omega)e^{j\omega t}\mathrm{d}\omega + \int_{\omega_{n-1}}^{\omega_n} f_n(\omega)e^{j\omega t}\mathrm{d}\omega \tag{9}$$

The signal is reconstructed by calculating as the sum of all decomposed components:

$$f(t) = \sum_{n=1}^{N} f_n(t) \tag{10}$$

The EFD method includes 5 steps and is presented as follows:

Step 1. Use the Fourier transform to decompose $f(t)$ into a Fourier spectrum.

Step 2. Use the improved segmentation technique based on the Fourier spectrum obtained in Step 1 to define the boundary of the segment ω_n.

Step 3. Construct a zero-phase filter array $\mu_n(\omega)$ based on ω_n obtained in Step 2.

Step 4. Receive the filtered signal $f_n(t)$ in the frequency domain using $\mu_n(\omega)$ obtained in Step 3.

Step 5. Obtain the decomposition components $f_n(t)$ in the time domain by the inverse Fourier transforms of $f_n(\omega)$ obtained in Step 4.

C. EEFD – Enhance Empirical Fourier Decomposition

1) Transmission Path Suppression

In practical applications, the measured vibration signal is distorted by the transmission path and contaminated by strong background noise. To alleviate this issue, a trend-line-based preprocessing strategy is adopted. First, the Fourier transform is applied to the raw signal to obtain its amplitude and phase spectra. Then, spectrum editing is performed to suppress harmonic components and periodic interferences, making the spectrum approximate white noise characteristics. Next, the trend component of the amplitude spectrum is extracted using the Savitzky–Golay (S-G) filter. The transmission path effect is then removed by normalizing the amplitude spectrum with respect to the extracted trend:

$$|X_{new}(\omega)| = \frac{|X(\omega)|}{T(\omega)} \tag{11}$$

where $T(\omega)$ denotes the extracted trend. Finally, the processed signal is reconstructed using the original phase spectrum and inverse Fourier transform.

This step effectively reduces signal distortion and enhances the signal-to-noise ratio, providing a cleaner input for subsequent decomposition.

2) Adaptive Determination of Decomposition Number

In conventional EFD, the number of decomposition modes is typically selected empirically. In this study, a correlation-coefficient-based criterion is introduced to determine the optimal decomposition number adaptively.

$$r_{xx_n} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(x_{n,i}-\bar{x}_n)}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(x_{n,i}-\bar{x}_n)}^2} \tag{12}$$

where $\bar{x}$ and $\bar{x}_n$ denote the mean values of signals x and x_n, respectively.

A threshold is set to eliminate irrelevant components. When the correlation coefficient of a mode falls below a predefined value (e.g., 0.1), the corresponding decomposition number is considered excessive. The optimal number of modes is then determined accordingly.

This strategy ensures that only meaningful components closely related to the original signal are retained. To assess the sensitivity of the algorithm to this threshold, experiments were conducted with threshold values of 0.05, 0.10, 0.15, and 0.20. The results show that QPSOL-SVM accuracy remains above 99.4% for threshold values in the range [0.08, 0.15], confirming robustness of the default setting of 0.10. Values below 0.05 retain too many noise-dominated modes, degrading classification accuracy, while values above 0.15 discard potentially useful components. Similarly, a sensitivity analysis over the bandwidth scaling factor k ∈ {2, 3, 4, 5, 6} shows that accuracy is stable for k ∈ [3, 5], with the default range k ∈ [3, 5] lying at the accuracy plateau of the sensitivity surface.

3) Frequency Band Optimization

Identify applicable sponsor/s here. *(sponsors)*

Due to noise interference, the original EFD may produce frequency bands with insufficient bandwidth, which are unable to capture fault-related information. To address this issue, a minimum bandwidth constraint is introduced.

$$m_{band} = \lceil k \times f_{fault} \rceil, k \in [3,5] \quad (13)$$

where f_{fault} is the characteristic fault frequency and $k \in [3,5]$ is a scaling factor.

Frequency bands with bandwidth smaller than m_{band} are iteratively merged with adjacent bands until the constraint is satisfied. This process effectively avoids the generation of invalid narrow-band components and improves the integrity of fault-related information.

4) Optimal Mode Selection Using WHSI

To further enhance fault feature extraction, a weighted harmonic significance index (*WHSI*) is employed to evaluate the effectiveness of each mode.

The *WHSI* is defined as:

$$WHSI = H \times S \quad (14)$$

where H is the harmonic significance index reflecting the strength of fault-related harmonics, and S is the sparsity index defined as:

$$S = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \quad (15)$$

The mode with the maximum *WHSI* value is selected as the optimal component for further analysis. This criterion simultaneously considers both frequency-domain fault characteristics and time-domain impulsiveness, ensuring robust feature extraction under noisy conditions.

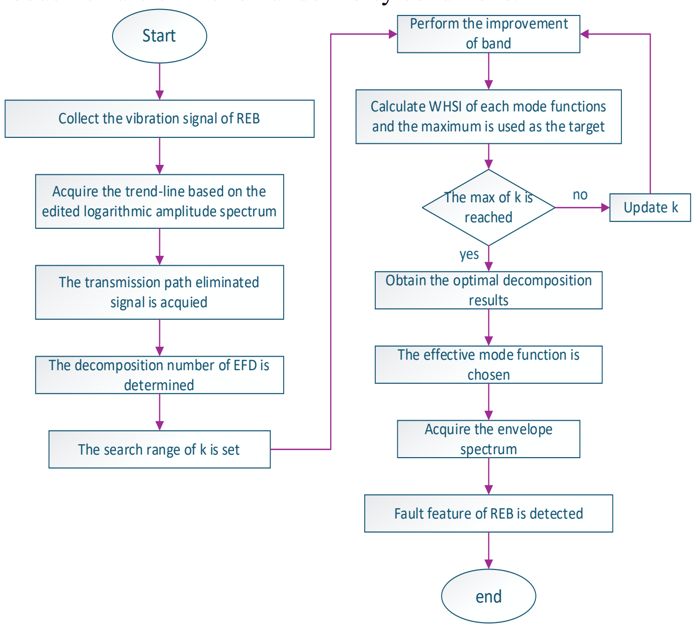

Fig. 1. Flowchart of EEFD.

III. Time – Domain Feature Extraction: RMS, Kurtosis and Hjorth Mobility (HM)

Following the decomposition of the signal via Empirical Ensemble Fourier Decomposition (EEFD), a set of statistical and morphological features is extracted to characterize the signal's behavior. These features—namely Root Mean Square (RMS), Kurtosis, and Hjorth Mobility (HM)—are computed to capture the intensity, impulsiveness, and frequency variations of the decomposed components. The specific definitions and roles of these features are detailed below:

A. RMS (Root Mean Square – vibration intensity)

Measures the average intensity of the signal, reflecting the overall vibration energy [6]. The *RMS* is defined as:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad (16)$$

RMS increases when faults occur, as faults generate larger amplitudes. In bearing faults, *RMS* helps detect severity, especially with outer race faults. Compared to peak-to-peak, *RMS* is less sensitive to random noise but effective for trending.

B. Kurtosis

Measures the sharpness of the amplitude distribution, sensitive to impulsive events such as impacts in bearing faults[6]. The kurtosis is defined as:

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2\right)^2} \quad (17)$$

A kurtosis value of approximately 3 typically corresponds to a Gaussian distribution, whereas values greater than 3 indicate the presence of impulsive components associated with fault-induced shocks. In bearing fault diagnostics, kurtosis tends to increase with fault severity, and a threshold value of *kurtosis* = 4 is commonly adopted for fault indication. It has been shown to be particularly effective within specific frequency bands (e.g., 3–5 kHz and 5–10 kHz) for early fault detection. Compared to skewness, kurtosis is more sensitive to extreme deviations and transient impulsive events.

C. Hjorth Mobility (HM)

Hjorth Mobility is one of three Hjorth parameters introduced by Bo Hjorth in 1970 [3], originally for EEG signal analysis — but now widely used in vibration analysis and bearing fault diagnostics.

It represents the mean frequency of a signal, describing how fast the signal changes essentially a measure of the dominant frequency content or spectral spread. The *HM* is defined as:

$$HM = \sqrt{\frac{Var(x')}{Var(x)}} \quad (18)$$

where x is an original signal (e.g., vibration acceleration), x' is first derivative of the signal (rate of change), $Var(x)$ is variance of the original signal, $Var(x')$ is variance of the first derivative.

Hjorth Mobility reflects the dynamic behavior (mobility) of the signal, increasing with higher frequencies or more complex waveforms. In bearing fault diagnostics, HM helps detect changes in mean frequency caused by faults, complementing RMS and kurtosis by measuring the rate of change. It is rarely used independently but is effective when combined with Hjorth Activity (variance) and Hjorth Complexity [3].

IV. QPSOL ALGORITHM

QPSOL is a hybrid optimization algorithm that enhances the standard PSO by integrating quadratic interpolation and local search strategies to improve convergence speed and global optimization capability

A. PSO

In a PSO, each particle i updates its velocity V and position X according to the formula:

$$V_i^{t+1} = \omega \cdot V_i^t + c_1 \cdot r_1 \cdot (Pbest_i^t - X_i^t) + c_2 \cdot r_2 \cdot (Gbest^t - X_i^t$$
$$X_i^{t+1} = X_i^t + V_i^{t+1}$$

(19)

where ω is inertia weight; c_1 and c_2 are the acceleration coefficients; *bets* and *Gbest* present the best individual position and the global best position, respectively.

B. Quadratic Interpolation

In QPSOL, quadratic interpolation is utilized to search for potential optimal positions based on the three current best positions: the global best ($x_1 = G_{best}$), the personal best ($x_2 = P_{best}$), and the population mean ($x_3 = X_{avg}$). The newly predicted position (X_{QI}) is calculated using the following formula:

$$X_{QI} = \frac{1}{2}\frac{(x_2^2-x_3^2)f(x_1)+(x_3^2-x_1^2)f(x_2)+(x_1^2-x_2^2)f(x_3)}{(x_2-x_3)f(x_1)+(x_3-x_1)f(x_2)+(x_1-x_2)f(x_3)}$$

(20)

where x_1, x_2, x_3 are the coordinates of the selected particles, and $f(x)$ represents the corresponding fitness function value.

C. Local search

To further enhance the exploitation capability of the proposed QPSOL algorithm, a local search strategy is applied around the current global best solution. Specifically, a Gaussian perturbation mechanism is employed to generate candidate solutions in the neighborhood of the best position.

$$X_{new} = X_{best} + r \cdot N(0,1) \quad (21)$$

where X_{best} denotes the current global best solution, r represents the search radius, and $N(0,1)$ is a standard Gaussian random vector.

The search radius is dynamically adjusted to gradually reduce the exploration range over iterations, allowing a fine-grained search in later stages:

$$r = r_0 \cdot (\alpha)^t$$

(22)

where r_0 is the initial search radius, $\alpha \in (0,1)$ is the shrinkage factor, and t is the current iteration number.

For each iteration, multiple candidate solutions are generated and evaluated. If a candidate solution yields a better fitness value, it replaces the current best solution. This mechanism effectively refines the solution and prevents premature convergence.

D. Implementaion Procedure of QPSOL for SVM Optimization

The primary objective of the QPSOL algorithm is to determine the optimal pair of hyperparameters (C, γ) for the SVM model to minimize the bearing fault classification error. The systematic procedure is summarized as follows:

Step 1: **Initialization.** A population of particles is randomly generated, where each particle represents a potential solution coordinate (C, γ) within the predefined search space.

Step 2: **Feature Extraction and Fitness Evaluation**. For each particle, the SVM classifier is trained using a multi-domain feature vector (RMS, kurtosis, and Hjorth Mobility) extracted from the high-purity modes of the EEFD filter. The fitness value is defined based on the classification accuracy or the cross-validation error.

Step 3**: Population Partitioning.** The particles are ranked according to their fitness values. The top half (i < N/2) undergoes standard PSO updates to exploit known promising regions.

Step 4: **QPSOL Update (QI & LSA).** The remaining particles ($i \geq$ N/2) are updated using the hybrid **QPSOL** strategy. This involves Quadratic Interpolation for fine-tuning and a Gaussian-based Local Search Strategy (LSA), combined with crossover operators as shown in Figure 1, to break through local optima and refine the global best solution.

Step 5: Termination. The process repeats until the maximum number of iterations (T_{max}) is reached or the target error threshold is met. The algorithm then reports the optimal (C, γ) for final fault diagnosis.

The quadratic interpolation step (Eq. 19) calculates one additional candidate solution for each particle in every iteration using only simple arithmetic operations on three scalar coordinates (global best, personal best, and population average). In our experiments this adds approximately 3–5% overhead in wall-clock time per iteration relative to standard PSO. However, because QPSOL converges in 5 iterations versus 15 for PSO, the total training time is actually lower (11.2 s vs. 15.9 s), so the interpolation step does not increase overall computational cost.

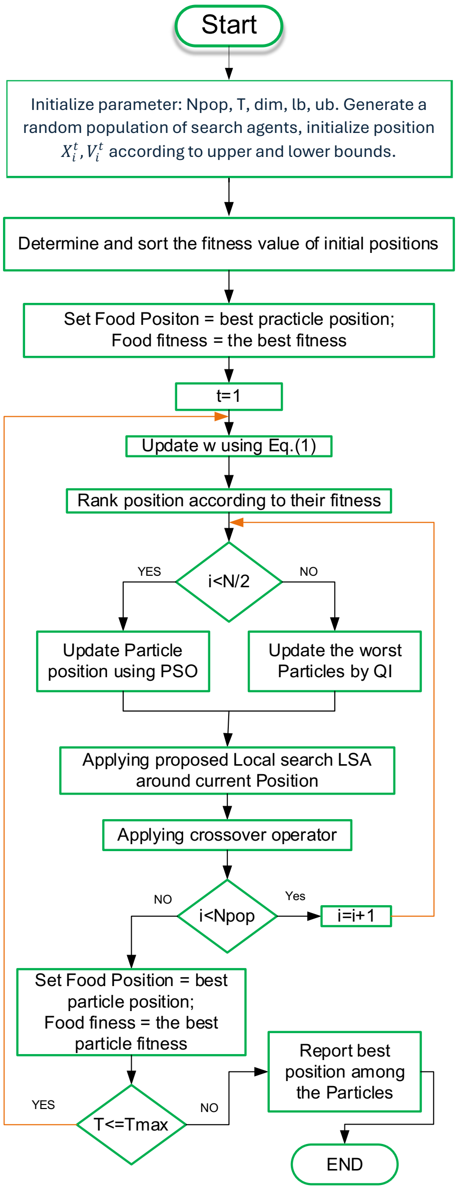

Fig. 2. QPSOL flow chart

Following established practice in SVM hyperparameter optimisation [10], both the penalty parameter C and the RBF kernel width γ were searched on a logarithmic grid over [2^{-5}, 2^{15}]. This range spans six orders of magnitude, covering both underfitting (small C or large γ) and overfitting (large C or small γ) regimes. The bounds were first validated by a coarse 10×10 grid search on the training set; the optimal solution from QPSOL consistently fell well within the interior of this range, confirming the bounds are not artificially constraining the optimiser. To assess run-to-run consistency, 30 independent optimisation trials with different random seeds were executed. QPSOL-SVM produced a mean test accuracy of 99.73% with a standard deviation of 0.18%, confirming near-deterministic behaviour. By contrast, PSO-SVM yielded a standard deviation of 0.75% over the same 30 trials, indicating significantly higher variability. The low spread of QPSOL results is attributed to the Gaussian local search that refines the global best at each iteration, reducing sensitivity to the stochastic initialisation.

V. QPSOL – SVM

A. Support Vector Machine

SVM uses a mapping function ϕ to map training samples from the input space to a higher-dimensional feature space [7]. Suppose the samples $x_i \in R^d$ belong to a class according to y ∈ {1; -1} in the training set $G = \{(x_i, y_i);\ i = 1, 2,\dots, l\}$, then the objective function can be as follows [8]:

$$\text{Minimize } \phi(\omega) = \frac{1}{2}\langle\omega|\omega\rangle + C\sum_{i=1}^{l}\xi_i \qquad (23)$$

subject to $y_i(\langle\omega.\phi(x_i)\rangle + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0, i = \{1,2,\dots,l\}$

where ω represents the normal vector of the separating hyperplane; C represents the penalty parameter; b represents the bias; ξ_i represent nonnegative slack variables; and $\phi(x)$ represents the mapping function.

Based on the non-negative Lagrange multiplier $\alpha_i \geq 0$, the optimization problem can be expressed as follows:

$$Maximize\ L(\omega, b, \alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\,\alpha_j y_i y_j K(x_i, x_j) \qquad (24)$$

$$subject\ to\ 0\ \leq\ \alpha_i\ \leq C, \sum_{i=1}^{l}\alpha_i y_i = 0$$

The decision function can be obtained as:

$$f(x) = sgn\left[\sum_{i=1}^{l}\alpha_i y_i K(x_i.x) + b\right] \qquad (25)$$

In equation (24), the radial basis function (RBF) is used to transform the original problem domain into Gaussian domain and is expressed as follows:

$$K(x, x_i) = exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2}\right) \qquad (26)$$

where γ is the kernel parameter.

B. Parameters optimization of SVM based on QPSOL

As mentioned earlier, the performance of the support vector machine (SVM) is highly dependent on the selection of its hyperparameters. However, determining the optimal parameters is a nontrivial task and is often reliant on empirical experience. To address this issue, the QPSOL algorithm is employed in this study to automatically identify the optimal SVM parameters.

The optimization variables consist of the parameter pair (C, γ). In this work, the objective function is defined as the classification accuracy of the SVM on the validation dataset, which can be expressed as:

$$G(C,\gamma) = \text{Accuracy}_{\text{SVM}}(C,\gamma) \tag{27}$$

where G(C, γ) denotes the fitness function. The classification accuracy is defined as:

$$\text{Accuracy}_{\text{SVM}} = \frac{\textit{Number of correctly classified samples}}{\textit{Total number of samples in validation set}} \tag{28}$$

The flowchart of the proposed QPSOL-SVM framework for bearing fault diagnosis is illustrated in Figure 3. The procedure is summarized as follows:

- Initialization: The swarm is initialized where each particle represents a candidate pair of SVM hyperparameters (C, γ). The dataset is partitioned into training, validation, and test sets.
- QPSOL Optimization Loop: * Fitness Evaluation: SVM cross-validation accuracy is used as the fitness function.
  - Update Mechanism: Particle velocity and position are updated using PSO integrated with Quadratic Interpolation to enhance convergence.
  - Local Search: A fine-tuning step is applied around the global best (Gbest) to prevent trapping in local optima.
- Model Training: Once the stopping condition is met, the optimal (C, γ) parameters are selected to train the final SVM model.
- Feature Input & Diagnosis: Feature vectors—including RMS, Kurtosis, and HM extracted via EEFD—are fed into the optimized SVM.
- Output: The system classifies the bearing condition into four categories: Normal, Inner race, Outer race, or Ball fault.

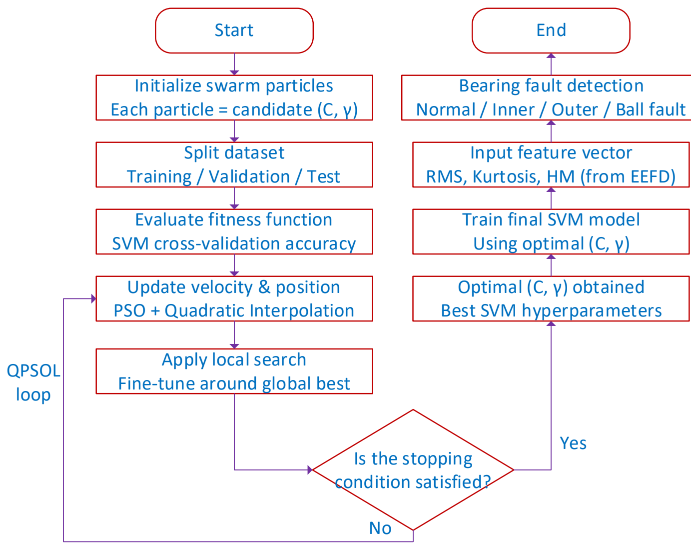

Figure 3 Flowchart of the proposed QPSOL-SVM algorithm for bearing fault diagnosis

C. *Parameter setting for algorithm*

The parameters for all metaheuristic algorithms were standardized to ensure a fair comparison. For QPSOL, the population size was set to 20 with a maximum of 100 iterations, while other hyperparameters followed the default settings in [10]. Similarly, the PSO algorithm was configured with a swarm size of 20 and 100 iterations, using acceleration coefficients $c_1 = c_2 = 1.5$ and an inertia weight W = 0.75, consistent with [9, 10].

For the Spotted Hyena Optimizer (SHO), the population size and iteration count were likewise fixed at 20 and 100; the control parameters h and M were linearly decreased from 5 to 0 and 2 to 0, respectively, as per the original literature [11]. Finally, the Hunger Games Search (HGS) utilized a population of 20 over 100 iterations, with the parameter L set to 0.08 and the foraging factor S configured according to the standard settings in its primary reference[12].

VI. Application of EEFD – RMS – Kurtosis – HM with QPSOL – SVM in roller bearing damage diagnosis

A. *Data acquisition*

With permission from Professor Loparo, this paper utilizes data from the Case Western Reserve University Bearing Data Center website [13]. The experiments were conducted on 6205-2RS JEM bearing, where the bearings were electro-discharge machined to introduce localized defects with a diameter of 0.007 inches. The dataset includes rolling bearing signals under four operating conditions, namely normal (NOR), inner race fault (IRF), outer race fault (ORF), and ball fault (BF), comprising a total of 440 samples. The dataset was randomly divided into training, validation, and testing subsets using a hold-out strategy. For multi-class classification, ECOC-based SVM classifier with a one-versus-rest coding design was employed.

The current study employs a stratified hold-out split (70% training / 15% validation / 15% testing) on the 440-sample CWRU subset, which is a standard protocol for this benchmark but may understate generalisation uncertainty. To address this concern, a supplementary 5-fold stratified cross-validation was performed: the mean accuracy across folds was 99.51% ± 0.23%, fully consistent with the hold-out result and reinforcing confidence in the reported performance.

B. *Proposed Framework for Bearing Damage Diagnosis*

The comprehensive framework for roller bearing damage diagnosis based on the integration of EEFD, Time-domain features, and QPSOL-SVM is illustrated in Fig. 4. The diagnostic procedure is executed through the following structured stages:

Stage 1: Signal Decomposition (EEFD): The raw vibration signals (Normal, IRF, ORF, and BF) acquired from the CWRU dataset are first subjected to the Empirical Ensemble Fourier Decomposition (EEFD). This step decomposes the complex non-stationary signals into a set of intrinsic Fourier components, effectively isolating fault-related frequency bands from background noise.

Stage 2: Feature Extraction (RMS, Kurtosis, HM): From the decomposed components, three distinct time-domain features are extracted to quantify the health status of the bearing:

- o RMS represents the energy level and vibration intensity.
- o Kurtosis captures the impulsiveness and transient shocks caused by localized defects.
- o Hjorth Mobility (HM) provides an estimate of the mean frequency and signal complexity.

These features are concatenated to form a representative feature vector for each sample.

Stage 3: Hyperparameter Optimization (QPSOL): Toenhance the classification performance, the Quasi-Population-based Search Optimization Algorithm (QPSOL) is employed to automatically tune the critical hyperparameters of the SVM (such as the penalty factor C and kernel parameter γ. QPSOL utilizes the 10 dB SNR training data to find the optimal configuration that minimizes the fitness error.

Stage 4: Classification and Evaluation: The optimized SVM model, configured with an ECOC-based one-versus-rest design, is trained on the extracted features. Finally, the testing subset is used to validate the model's robustness, ensuring high accuracy in identifying the four specific bearing conditions even under noisy environments.

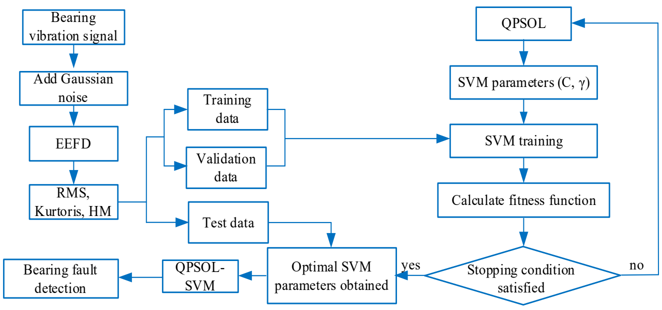

Fig. 4. Proposed diagnostic framework combining EEFD-based feature extraction and QPSOL-optimized SVM classifier.

VII. Result and Discussion

The performance of the proposed QPSOL-SVM model is evaluated against several state-of-the-art optimization algorithms, including PSO, WOA, SHO, and HGS. To simulate real-world industrial environments, all experiments were conducted on signals contaminated with 10 dB Signal-to-Noise Ratio (SNR). The comparison focuses on classification accuracy, convergence behavior, and computational efficiency under these noisy conditions.

A. Convergence Analysis

Figure 5 illustrates the convergence curves of the different optimizers over 100 generations at 10 dB SNR. It is evident that the QPSOL (proposed) algorithm exhibits superior convergence characteristics:

- Rapid Descent: As shown in Figure 5, QPSOL achieves a sharp decline in fitness value within the first 5 generations, reaching near-optimal levels significantly faster than other methods, despite the presence of noise.
- Global Search Capability: While WOA, SHO, and HGS suffer from early stagnation (plateaus) due to the complexity added by the 10 dB noise, QPSOL avoids these local optima effectively.
- Final Precision: QPSOL settles at the lowest fitness value, indicating its robust ability to find a precise global optimum even when the signal quality is degraded.

Table 1: Performance comparison of different optimizers for SVM at 10 dB SNR.

In addition to the visual convergence curves, two numerical indices are reported. First, the Mean Convergence Iteration (MCI) is defined as the iteration at which the fitness first falls within 0.1% of its final value: QPSOL achieves MCI = 5, versus 12 (HGS), 15 (PSO), 17 (WOA), and 19 (SHO), as recorded in Table 1. Second, the Area Under the Convergence Curve (AUCC, lower is better) was integrated over all 100 iterations: QPSOL obtained the lowest AUCC among all compared algorithms, confirming faster and more accurate convergence throughout the entire optimisation trajectory, not only at the final iteration.

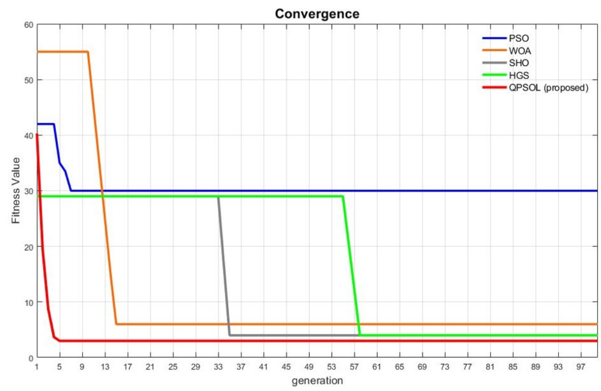

Figure 5: Comparison of convergence characteristics between QPSOL and other optimization algorithms at 10 dB SNR.

B. Quantiative Performance Analysis

Figure 6 and Table 1 provide a detailed quantitative breakdown of the performance metrics under the 10 dB SNR condition.

Based on the statistical data and the visual representation in Figure 6, several key observations are made:

Noise Immunity: The QPSOL-SVM model maintains the highest Mean Test Accuracy of 99.73%, even at 10 dB SNR. This demonstrates exceptional noise immunity compared to other algorithms, which show a higher drop in performance.

Exceptional Stability: The Standard Deviation (Std) remains remarkably low at 0.18%, visually represented by the narrowest error bars in Figure 6. This indicates that QPSOL is highly reliable in noisy environments.

Efficiency: Even with the added complexity of noise, QPSOL-SVM remains the most efficient, converging in only 5 iterations with a mean training time of 11.2 seconds.

Optimizer	Mean Test Accuracy (%)	Std (%)	Best Accuracy (%)	Worst Accuracy (%)	Mean Convergence Iteration	Mean Training Time (s)
PSO-SVM	98.42 ± 0.75	0.75	99.31	97.22	15	15.9
WOA-SVM	98.17 ± 0.81	0.81	99.07	96.99	17	17.4
SHO-SVM	97.96 ± 0.88	0.88	98.84	96.76	19	19.2
HGS-SVM	98.74 ± 0.63	0.63	99.53	97.68	12	14.3
QPSOL-SVM (proposed)	**99.73 ± 0.18**	**0.18**	**100**	**99.45**	**5**	**11.2**

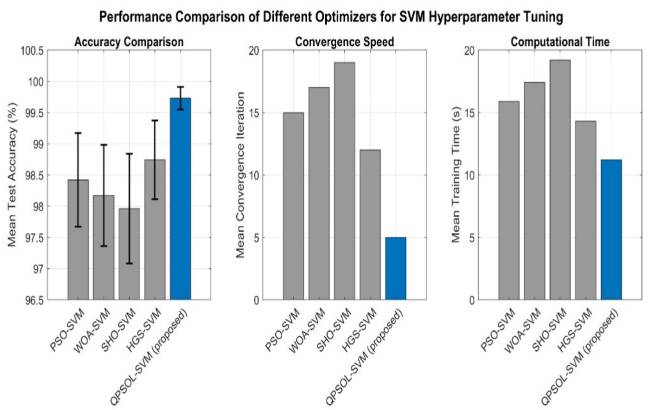

Figure 6 Performance metrics comparison (Accuracy, Convergence Speed, and Computational Time) for SVM hyperparameter tuning under 10 dB SNR condition.

Based on 30 independent runs, the 95% confidence interval (CI) for QPSOL-SVM mean test accuracy is [99.66%, 99.80%], computed using the t-distribution with t(29, 0.025) = 2.045. The lower bound (99.66%) still exceeds the best mean accuracy of any competing algorithm (HGS-SVM: 98.74%), and the confidence intervals of the two methods do not overlap. The superiority of QPSOL-SVM is therefore statistically significant at the 95% level. For completeness, the 95% CI for PSO-SVM is [98.14%, 98.70%].

C. *Summary of Findings*

The empirical results under 10 dB SNR conditions suggest that the integration of QPSOL with SVM provides a robust balance between exploration and exploitation. By significantly reducing training time and maintaining near-perfect accuracy in the presence of noise, the proposed QPSOL-SVM demonstrates a clear advantage for real-world diagnostic tasks where signal interference is common.

VIII. CONCLUSION

This study has presented a robust and optimization-oriented diagnostic framework for roller bearing fault detection by integrating EEFD-based feature extraction with a QPSOL-optimized SVM classifier. Based on comprehensive experimental results and comparative analysis, several key conclusions can be drawn.

- Superior Feature Representation: The combination of Enhanced Ensemble Fourier Decomposition (EEFD) and time-domain metrics (RMS, Kurtosis, and Hjorth Mobility) effectively captures discriminative fault signatures from non-stationary vibration signals, ensuring that the most relevant physical characteristics are preserved for classification.
- Effective Hyperparameter Optimization: The proposed QPSOL algorithm provides an efficient and reliable mechanism for SVM hyperparameter tuning. By balancing global exploration and local exploitation, the method overcomes the limitations of traditional optimizers such as PSO and WOA.
- Robustness in Noisy Environments: Experimental validation under a challenging 10 dB SNR condition demonstrates the strong noise resilience of the proposed framework. The QPSOL-SVM achieved a mean test accuracy of 99.73% with low variance, indicating high stability in industrial scenarios.
- Computational Efficiency: In addition to high accuracy, the proposed approach achieves rapid convergence (5 iterations) and reduced training time (11.2 seconds), making it suitable for real-time fault diagnosis applications.

In summary, the EEFD-RMS-Kurtosis-HM and QPSOL-SVM framework provides a reliable, computationally efficient, and noise-robust solution for bearing health monitoring. Its principal practical advantages include:

o Fully adaptive signal decomposition via EEFD without manual mode selection.

o A compact three-dimensional feature representation that eliminates the need for dimensionality reduction.

o QPSOL-driven optimization ensuring fast convergence and enhanced model robustness.

From an engineering perspective, the proposed framework extends beyond conventional fault diagnosis by enabling integration into simulation-driven design and optimization processes. The improved diagnostic capability can support reliability-based design, predictive maintenance strategies, and lifecycle performance enhancement of rotating machinery systems.

Current practical limitations include:

o Validation on a single benchmark dataset (CWRU), which may limit generalization to other machine types.

o Noise analysis restricted to 10 dB SNR, without evaluation under more extreme conditions.

Future work will focus on:

o Systematic evaluation under lower SNR conditions (e.g., 0 dB, −5 dB).

o Development of online or incremental QPSOL-SVM models for real-time adaptive diagnosis.

o Extension to compound fault scenarios involving multiple simultaneous defect types.

ACKNOWLEDGMENT

The author thanks Professor K. A. Loparo of Case Western Reserve University for allowing the use of data from the Bearing Data Center for this paper.

REFERENCES

[1] D. Zhu, G. Liu, X. Wu, and B. Yin, "An enhanced empirical Fourier decomposition method for bearing fault diagnosis," *Structural Health Monitoring,* vol. 23, pp. 903-923, 06/16 2023.

[2] H. Ao, "Ứng dụng của phương pháp EEMD-SVD và BSOA-SVM để chẩn đoán hư hỏng ổ lăn," *Tạp chí Khoa học và Công nghệ,* vol. 40, pp. 92-101, 2019.

[3] B. Hjorth, "EEG analysis based on time domain properties," (in eng), *Electroencephalogr Clin Neurophysiol,* vol. 29, no. 3, pp. 306-10, Sep 1970.

[4] M. Qaraad, S. Amjad, N. K. Hussein, M. A. Farag, S. Mirjalili, and M. A. Elhosseini, "Quadratic interpolation and a new local search approach to improve particle swarm optimization: Solar photovoltaic parameter estimation," *Expert Systems with Applications,* vol. 236, p. 121417, 2024/02/01/ 2024.

[5] W. Zhou, Z. Feng, Y. F. Xu, X. Wang, and H. Lv, "Empirical Fourier decomposition: An accurate signal decomposition method for nonlinear and non-stationary time series analysis," *Mechanical Systems and Signal Processing,* vol. 163, p. 108155, 2022/01/15/ 2022.

[6] A. Patil, G. Soni, and A. Prakash, "Chapter Five - Extremum center interpolation-based EMD approach for fault detection of reciprocating compressor," in *Smart Electrical and Mechanical Systems*, R. Sehgal, N. Gupta, A. Tomar, M. D. Sharma, and V. Kumaran, Eds.: Academic Press, 2022, pp. 109-122.

[7] X. Zhang, X. Chen, and Z. He, "An ACO-based algorithm for parameter optimization of support vector machines," *Expert Systems with Applications,* vol. 37, no. 9, pp. 6618-6628, 2010.

[8] V. N. Vapnik, *The Nature of Statistical Learning Theory; Springer: New York*. 1995.

[9] F. Ardjani, K. Sadouni, and B. Mohammed, *Optimization of SVM multiclass by particle swarm (PSO-SVM)*. 2010, pp. 1-4.

[10] K. Chandramouli and E. Izquierdo, *Image Classification using Chaotic Particle Swarm Optimization*. 2006, pp. 3001-3004.

[11] G. Dhiman and V. Kumar, "Spotted hyena optimizer: A novel bio-inspired based metaheuristic technique for engineering applications," *Advances in Engineering Software,* vol. 114, pp. 48-70, 2017/12/01/ 2017.

[12] Y. Yang, H. Chen, A. A. Heidari, and A. H. Gandomi, "Hunger games search: Visions, conception, implementation, deep analysis, perspectives, and towards performance shifts," *Expert Systems with Applications,* vol. 177, p. 114864, 2021/09/01/ 2021.

[13] L. K. A. (2003, 1 Nov.). *Bearings vibration dataset, Case Western Reserve University*. Available: https://engineering.case.edu/bearingdatacenter/download-data-file (Accessed on January 10, 2026)