

Article

A Hybrid Machine Learning Framework for Early Fault Detection in Power Transformers Using PSO and DMO Algorithms

Mohammed Alenezi ^{1,*}, Fatih Anayi ¹, Michael Packianather ²  and Mokhtar Shouran ³ ¹ Wolfson Centre for Magnetics, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK² High-Value Manufacturing Group, School of Engineering, Cardiff University, Cardiff CF24 3AA, UK³ Libyan Center for Engineering Research and Information Technology, Bani Walid 00218, Libya

* Correspondence: alenezim1@cardiff.ac.uk

Abstract: The early detection of faults in power transformers is crucial for ensuring operational reliability and minimizing system disruptions. This study introduces a novel machine learning framework that integrates Particle Swarm Optimization (PSO) and Dwarf Mongoose Optimization (DMO) algorithms for feature selection and hyperparameter tuning, combined with advanced classifiers such as Decision Trees (DT), Random Forests (RF), and Support Vector Machines (SVM). A 5-fold cross-validation approach was employed to ensure a robust performance evaluation. Feature extraction was performed using both Discrete Wavelet Decomposition (DWD) and Matching Pursuit (MP), providing a comprehensive representation of the dataset comprising 2400 samples and 41 extracted features. Experimental validation demonstrated the efficacy of the proposed framework. The PSO-optimized RF model achieved the highest accuracy of 97.71%, with a precision of 98.02% and an F1 score of 98.63%, followed by the PSO-DT model with a 95.00% accuracy. Similarly, the DMO-optimized RF model recorded an accuracy of 98.33%, with a precision of 98.80% and an F1 score of 99.04%, outperforming other DMO-based classifiers. This novel framework demonstrates significant advancements in transformer protection by enabling accurate and early fault detection, thereby enhancing the reliability and safety of power systems.

Keywords: power transformer protection; fault detection; machine learning classifiers; early fault diagnosis; optimization algorithms; feature extraction



Academic Editor: Fengshou Gu

Received: 3 March 2025

Revised: 24 March 2025

Accepted: 12 April 2025

Published: 15 April 2025

Citation: Alenezi, M.; Anayi, F.; Packianather, M.; Shouran, M. A Hybrid Machine Learning Framework for Early Fault Detection in Power Transformers Using PSO and DMO Algorithms. *Energies* **2025**, *18*, 2024. <https://doi.org/10.3390/en18082024>

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

1. Introduction

Power transformers are vital components in electrical power networks, responsible for voltage regulation and power transmission [1]. Transformer failures can lead to severe financial losses, operational disruptions, and system instability [2]. Therefore, accurate and early fault detection is crucial for maintaining system reliability and preventing catastrophic failures [3,4].

Traditional fault detection methods, such as differential protection schemes and dissolved gas analysis (DGA), have limitations in distinguishing between internal and external faults, leading to false positives or missed detections [5]. Recent advancements in machine learning (ML) have significantly improved fault classification [6], but their performance is highly dependent on effective feature selection and hyperparameter tuning.

In response to the limitations of conventional diagnostic methods, machine learning (ML) [7] techniques have emerged as a reliable approach to enhance the accuracy of transformer fault identification and classification. Among supervised algorithms, Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) models have been

widely recognized for their effectiveness in this domain. DTs function by systematically dividing the input data into smaller subsets based on feature-specific thresholds, resulting in a hierarchical and interpretable structure ideal for classification tasks. Despite their computational efficiency, DT models are known to overfit, especially when applied to high-dimensional datasets [8]. To counteract this, Random Forest (RF) [9] extends the DT methodology by constructing an ensemble of decision trees and combining their outputs, which significantly enhances stability and predictive accuracy. This ensemble strategy improves generalization capabilities, making RF particularly advantageous for analyzing large-scale and complex data scenarios [10,11].

Metaheuristic optimization techniques have emerged as powerful tools to enhance ML models by improving feature selection and classifier performance. Among them, Genetic Algorithm (GA) [12], Grey Wolf Optimizer (GWO) [13], and Differential Evolution (DE) [14] have been widely used. However, these methods often suffer from premature convergence and suboptimal search efficiency. In contrast, Particle Swarm Optimization (PSO) and Dwarf Mongoose Optimization (DMO) offer superior exploration–exploitation trade-offs, leading to a more reliable and efficient tuning of ML models [15,16].

This study proposes a hybrid optimization framework combining PSO and DMO to enhance ML-based fault detection in power transformers. The framework integrates Discrete Wavelet Transform (DWT) and Matching Pursuit (MP) for feature extraction, followed by PSO and DMO for selecting the most relevant features and optimizing classifier parameters. This results in a robust system capable of detecting faults with a high accuracy and generalizability.

Research Contributions

This study introduces significant theoretical and practical contributions:

- **Hybrid Metaheuristic Optimization:** The integration of PSO and DMO improves feature selection and classifier hyperparameter tuning, ensuring a robust classification.
- **Computational Efficiency:** The proposed DMO-based approach converges faster than conventional PSO, reducing computational overhead.
- **Interpretability vs. Deep Learning:** Unlike black-box CNN-based methods, our approach is interpretable and requires less computational power.
- **Real-World Deployment:** The proposed method is adaptable for SCADA-based fault detection systems, making it suitable for online monitoring.

2. Related Work

A thorough literature review has been updated to include studies from 2020–2024, ensuring alignment with modern power transformer fault detection techniques. Additionally, Table 1 highlights key differences between the authors' work and recent CNN-based, hybrid ML, and bio-inspired methods.

Table 1. A comparison between CNN-based, Hybrid ML, and bio-inspired methods.

Methodology	Feature Selection	Optimization Technique	Classifier Used	Performance	Ref.
CNN-Based	Manual Feature Extraction	None	CNN	92.5% Accuracy	[17]
Hybrid ML (DWT + ML)	PCA	GA	SVM	94.1% Accuracy	[18]
Bio-Inspired Approach	Wrapper-Based	GWO	RF	96.3% Accuracy	[19]
DWT + ML	PSO-DMO Hybrid Selection	PSO + DMO	RF + SVM + DT	98.33% Accuracy	This Study

Comparison with Deep Learning-Based Methods

Recent studies have explored deep learning-based fault prognosis, such as CNNs and hybrid LSTM models. However, these models require large, labeled datasets and are computationally expensive. In contrast, the PSO-DMO hybrid approach offers the following (Table 2):

- A lower computational complexity than CNN-based methods.
- A better generalization for small to medium-sized datasets.
- Explainability in terms of selected features and classifier decisions.

Table 2. A comparison between recent studies.

Method	Feature Extraction	Optimization	Model Complexity	Accuracy	Ref.
CNN-Based	Raw Signal Processing	None	High	96.5%	[20]
LSTM-Based	Time-Series Analysis	None	Very High	97.2%	[21]
Authors' Approach	DWT + MP	PSO + DMO	Medium	98.33%	This Study

3. Materials and Methods

3.1. Experimental Setup

To validate our proposed methodology, a three-phase power transformer experimental testbed was constructed in the laboratory. The setup consists of the following:

- Data Acquisition (DAQ) System: Used for real-time measurement and data recording.
- Current Transformers (CTs): For fault current measurement.
- Switch Relays and Protection Resistance: Implemented for fault simulation.
- LabVIEW Interface: Used for real-time monitoring and data visualization.

The proposed method was validated using a three-phase power transformer experimental testbed. The setup consists of the following:

- Transformer Specifications: A 5 kVA, 416 V/240 V three-phase transformer was used for fault simulations.
- Fault Simulation Unit: A controlled switching mechanism was employed to introduce faults, including line-to-ground, line-to-line, and turn-to-turn faults.
- Data Acquisition System (DAQ): A NI cDAQ-9178 system with an NI-9205 analog input module was used for real-time current and voltage measurement.
- Current and Voltage Sensors: High-precision LEM LA 55-P current sensors and LV 25-P voltage sensors were utilized for accurate signal acquisition.
- Software Integration: The setup was interfaced with LabVIEW for real-time monitoring, signal logging, and data processing.

3.2. Dataset Description

The dataset consists of 600 samples across 6 fault categories, including a healthy transformer class. Feature extraction was performed using Discrete Wavelet Transform (DWT) and Matching Pursuit (MP), generating 41 extracted features. A 5-fold cross-validation approach was used to ensure a robust performance evaluation.

Figure 1 illustrates the complete experimental setup, including the DAQ system, CTs, relays, and control interface.

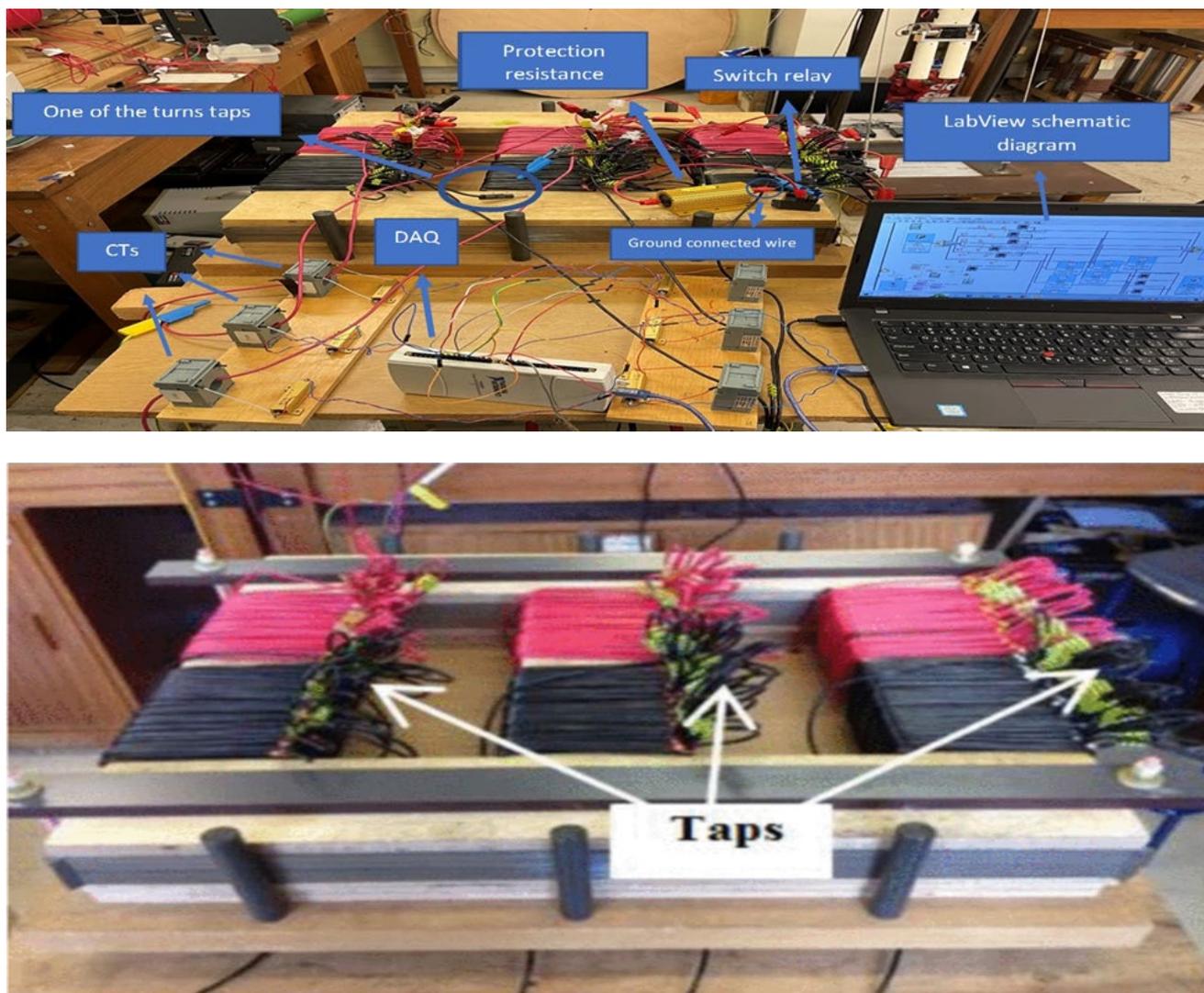


Figure 1. A diagram of the test system used in the laboratory.

3.3. Fault Simulation and Data Collection

To assess the performance of the authors' proposed method, the authors simulated common transformer faults, summarized in Table 3.

Table 3. The performance of the proposed method.

Fault Type	Description
Single Line-to-Ground (SLG) Fault	A single phase contacts the ground, leading to an unbalanced current.
Line-to-Line (LL) Fault	A short circuit between two phases, increasing current flow and heating.
Turn-to-Ground (TG) Fault	A winding turn contacts the ground, affecting insulation integrity.
Turn-to-Turn (TT) Fault	Short circuit within the same winding, leading to localized heating and damage.
External Faults	Faults outside the transformer affecting secondary protection mechanisms.

Fault events were induced using switch relays, and LabVIEW captured real-time signal variations. Data was collected at a 1 kHz sampling rate, then filtered using a Butterworth low-pass filter (500 Hz cutoff) to remove noise.

3.4. Data Processing and Feature Extraction

The collected signals were decomposed using Discrete Wavelet Transform (DWT) to extract time–frequency domain features. Feature selection was performed using PSO and DMO, ensuring an optimal classifier performance.

4. Research Methodology

This stage is structured into three primary components: extracting features, selecting the most relevant ones, and applying classification algorithms.

4.1. Feature Extraction

This paper utilizes two effective signal processing methods—Discrete Wavelet Transform (DWT) and Matching Pursuit (MP)—for feature extraction. These techniques were implemented using the Wavelet Analyzer Toolbox in MATLAB software (version R2022a), specifically employing the 1D Matching Pursuit tool for MP and the one-dimensional wavelet tool for DWT.

- Time–frequency characteristics are derived from the acquired signals using Discrete Wavelet Transform (DWT) and Matching Pursuit (MP) techniques.
- These techniques enhance the detection of transient fault characteristics that are often missed by traditional feature extraction methods.

4.1.1. Discrete Wavelet Transform (DWT)

Wavelet-based approaches are highly effective for transforming raw signals into a combined time–frequency representation, enabling a simultaneous analysis of both domains. Widely adopted across various engineering disciplines, Wavelet Transform (WT) has become a trusted method for identifying and diagnosing faults in mechanical systems. Initially proposed by Morlet in 1984, WT was developed to overcome the resolution limitations inherent in classical methods such as the Short-Time Fourier Transform (STFT), providing deeper insights into subtle signal characteristics. Unlike the Fast Fourier Transform (FFT) and related techniques, WT enables multiscale analysis by dynamically shifting and scaling wavelets across time and frequency. This capability allows for effective noise reduction without compromising signal integrity and ensures the precise localization of signal features in both domains [22,23].

The Discrete Wavelet Transform (DWT), a specific implementation of WT, is extensively used for analyzing various signals, including electromyography, thermal images, current, and vibration data. DWT operates by discretizing both the scale and time parameters, as described below:

$$WTx(t, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} vx(\tau) \psi\left(\frac{(t - \tau)}{a}\right) d\tau \quad (1)$$

where $\psi\left(\frac{(t - \tau)}{a}\right)$ is the wavelet obtained by scaling and translating the wavelet basis, $\psi(t)$ represents the scale parameter, t is the time shift, and $\frac{1}{\sqrt{a}}$ is a normalization factor. Alternatively, DWT can also be defined as follows:

$$W(m, n, \Psi) = a_0^{-\frac{m}{2}} \int x(t) \Psi * (a_0^{-m} t - nb_0) dt \quad (2)$$

where $x(t)$ is the input signal, m and n are integers, and the scale a and time b are given by the following:

$$a = a_0^m, \quad b = na_0^{mb_0} \quad (3)$$

4.1.2. Multiresolution Analysis

DWT decomposes a signal into multiresolution coefficients to extract useful features for analysis. This decomposition process uses low-pass and high-pass filters to split the signal into components across different frequency ranges:

- High-frequency components are referred to as detail coefficients (D).
- Low-frequency components are referred to as approximation coefficients (A), which provide a better frequency resolution.

The decomposition is expressed mathematically as follows:

$$Z_{high} = \sum x[k] \cdot g[2k - n] \quad (4)$$

$$Z_{low} = \sum x[k] \cdot h[2k - n] \quad (5)$$

Wavelet decomposition extracts approximation coefficients (Z_{low}) and detail coefficients (Z_{high}) from the input signal $x[k]$ using a low-pass filter ($g[2k - n]$) and a high-pass filter ($h[2k - n]$).

The maximum decomposition level is selected to ensure at least one coefficient remains unaffected by edge effects from signal extension, as defined in Equation (6):

$$Max_level = \left\lceil \log_2 \frac{data_len}{filter_len - 1} \right\rceil \quad (6)$$

when the $filter_len$ refers to the length of the filter or the wavelet object.

In summary, DWT's ability to decompose a signal into approximation and detail coefficients makes it a versatile tool for feature extraction, fault detection, and time–frequency analysis.

4.1.3. Decomposition Using Discrete Wavelet Transform (DWT)

DWT is widely recognized as a reliable method for extracting significant features in power transformer fault analysis. In contrast to conventional Fourier-based techniques, DWT offers a time–frequency perspective, which is particularly useful for identifying transient and localized fault behaviors. The decomposition process involves recursively analyzing the approximation components of the signal at each level, continuing this procedure until the specified resolution depth is achieved [24].

This study utilizes five mother wavelets (db7, sym3, coif4, bior6.8, and rbior6.8) in MATLAB's Wavelet Toolbox to analyze internal and external faults. A six-level wavelet decomposition extracts key frequency components, refining the signal into approximation and detail coefficients.

A total of 40 time–frequency features were extracted for each wavelet, resulting in a dataset of 2400 samples and 40 attributes, forming the basis for fault classification. Figure 2 illustrates the six-level decomposition of a power transformer signal under an internal fault, highlighting how DWT isolates high-frequency transient features, reinforcing its effectiveness in fault detection and diagnosis.

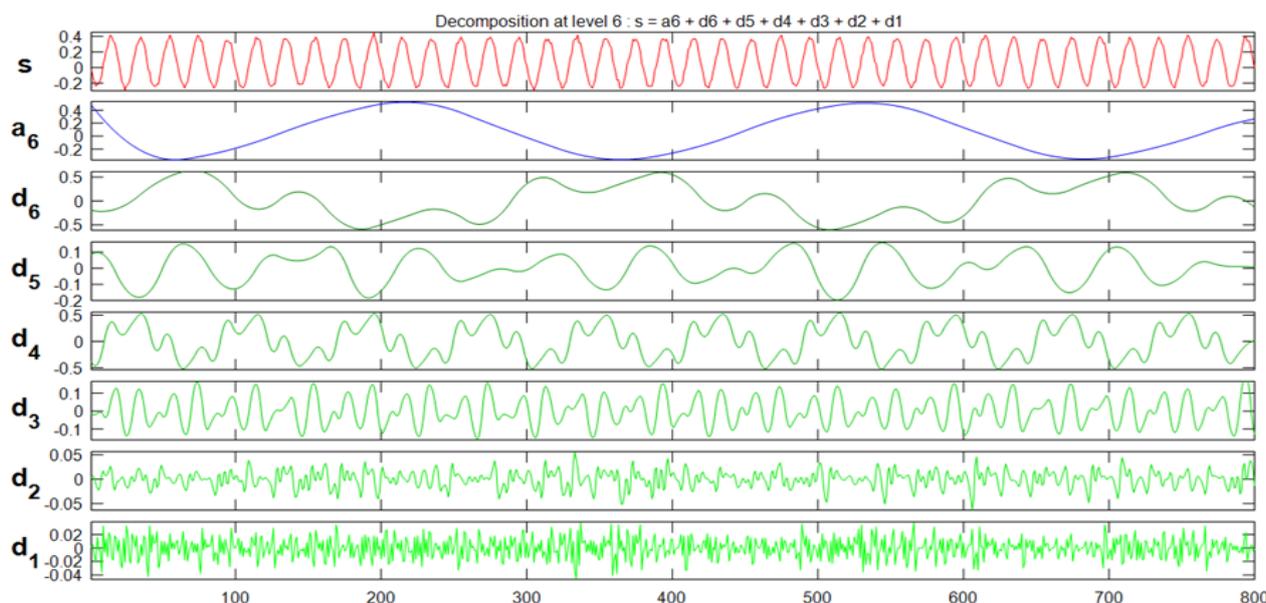


Figure 2. Discrete Wavelet Transform (DWT).

Justification for DWT Selection

DWT was selected over alternatives such as Continuous Wavelet Transform (CWT) and Stationary Wavelet Transform (SWT) due to the following:

- A lower computational cost than CWT.
- A more concise feature representation than SWT.
- Proven success in transformer signal processing tasks.

4.1.4. Matching Pursuit (MP)

MP is a multiscale signal analysis method that utilizes an overcomplete dictionary to iteratively break down signals. It represents the input signal as a linear combination of basic waveforms, referred to as atoms, which are selected from a redundant set of predefined functions. These atoms are chosen based on their ability to closely replicate the original signal, ensuring a detailed approximation within the time–frequency space.

4.1.5. Implementation of the Orthogonal Matching Pursuit Algorithm

In this study, the (OMP) PursXuit (OMP) algorithm [25] was implemented using five distinct signal components from the OMP dictionary, as detailed in Table 4. The collected signal data was analyzed under both healthy and faulty conditions utilizing MATLAB's OMP functionality. The key parameters were carefully configured, setting the maximum iteration limit to 500 and the maximum relative error based on the L1 norm to 0.01%. These configurations ensure efficient signal decomposition and precise feature extraction.

Table 4. (OMP) components [26].

Dictionary Parameter	Description
sym4-lev5	A five-level symmetric wavelet transform characterized by four vanishing moments was applied.
wpsym4-lev5	Wavelet packet transform using symmetric wavelets with five levels and four vanishing moments
Dct	Discrete Cosine Transform

Table 4. Cont.

Dictionary Parameter	Description
Sin	Sine-based sub-dictionary
Cos	Cosine-based sub-dictionary

Using (OMP), eight statistical metrics were derived as follows: mean, median, standard deviation, median absolute deviation, mean absolute deviation, L1 norm, L2 norm, and maximum norm, as outlined in the subsequent equations.

$$\text{Mean } \mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (7)$$

$$\text{Median, Med} = \frac{1}{2} \left(\times \left(\left[\frac{N+1}{2} \right] \right) + \times \left(\left[\frac{N}{2} \right] 1 \right) \right) \quad (8)$$

$$\text{standard deviation, } \sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1}} \quad (9)$$

$$\text{Median Absolute Deviation, med_AD} = \text{median}(|x_i - \text{median}(X)|) \quad (10)$$

$$\text{Mean Absolute Deviation, } \sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N-1} \quad (11)$$

$$\text{L1 norm } \|L\|_1 = \sum_{i=1}^N |x_i| \quad (12)$$

$$\text{L2 norm } \|L\|_2 = \sqrt{\sum_{i=1}^N |x_i|^2} \quad (13)$$

$$\text{Max norm } \|L\|_\infty = \max\{|x_i|\} \quad (14)$$

The acquired signal, denoted as x_i where $i = 1, 2, 3, \dots, N$, was analyzed using different mathematical norms. The L_1 **norm** represents the sum of the absolute values of all components, the L_2 **norm** is the square root of the sum of the squared absolute values, and the **LInf infinity norm** identifies the maximum absolute value among the components. These norms provide a quantitative measure of error and approximation accuracy in signal reconstruction.

Figure 3A shows the signal and its approximation after 500 iterations of OMP, achieving 100.16% retained energy. The signal exhibits oscillatory behavior with amplitudes ranging from -0.2 to 0.4 . The low relative error values ($L_2 = 0.21\%$, $L_{\infty} = 0.39\%$, and $L_1 = 0.18\%$) confirm the accuracy of the approximation.

Figure 3B illustrates the distribution of 500 selected coefficients across different dictionary components from a total of 3200 available coefficients. For instance, 131 out of 800 coefficients were chosen in the sym4—lev5 wavelet dictionary, while the sin and cos components retained 54 and 69 coefficients out of 400, respectively. These selections highlight the sparsity and significance of the chosen coefficients across transformation domains.

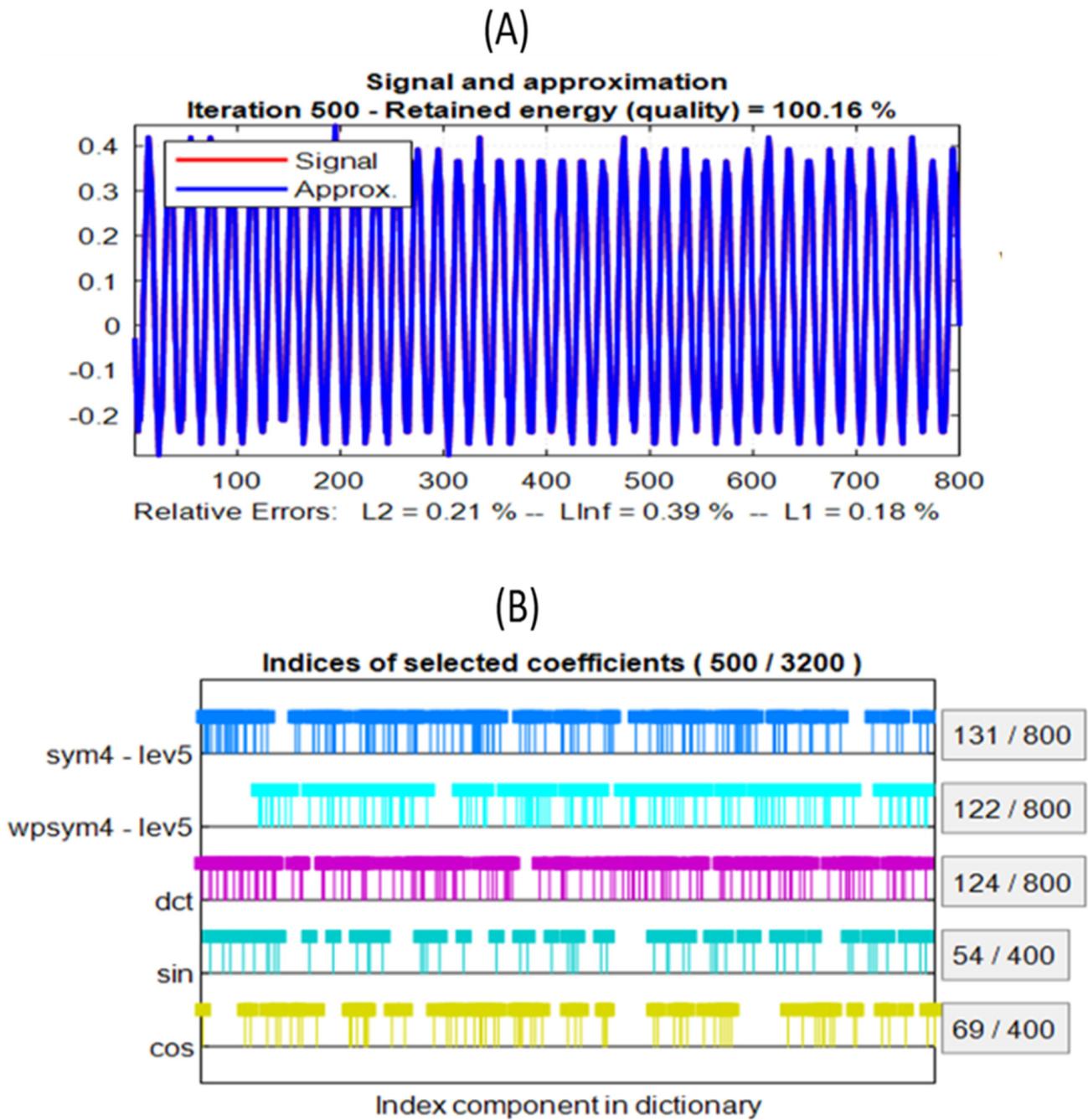


Figure 3. The selected coefficients and signals.

Figure 4 shows the decomposition of the original signal into different transformation bases, including wavelet (sym4—lev5, and wpsym4—lev5), discrete cosine transform (DCT), sinusoidal (sin), and cosine (cos) components. The *y*-axis represents signal amplitude, ranging from -0.3 to 0.4 . The original signal (red) has the highest amplitude, while its decomposed components capture different frequency bands and structural features. Wavelet components highlight localized details, DCT provides a global frequency view, and sin/cos components reflect periodic structures. Together, these decompositions preserve key features in a sparse representation, enabling an efficient signal reconstruction.

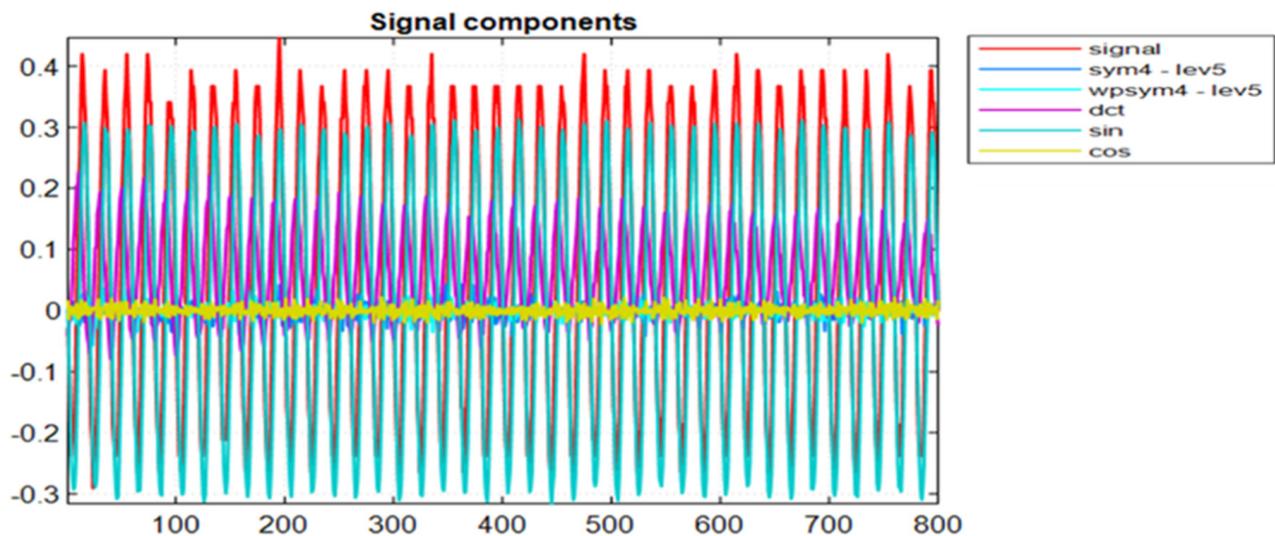


Figure 4. Reconstructed signal components.

Figure 5 shows the residual signal, representing the difference between the original and reconstructed signals. The y -axis indicates residual values, ranging from 0 to 1.5, where larger values reflect reconstruction errors. While the decomposition captures the most signal characteristics, some localized variations remain unaccounted for, suggesting areas for further refinement in signal representation.

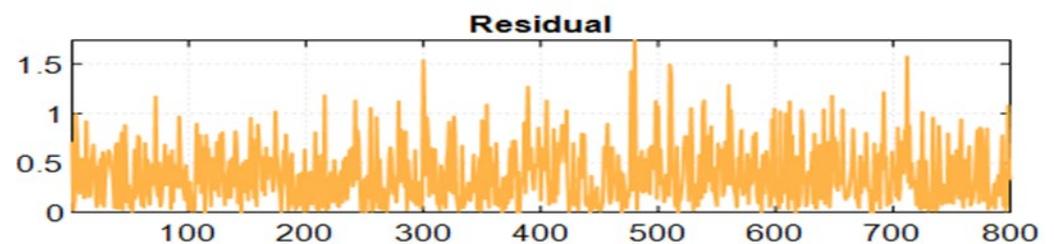


Figure 5. The residual signal.

To further validate the OMP's performance, an internal fault scenario is demonstrated in Figures 3–5, utilizing 800 sample points from the signal. This analysis substantiates the claim that OMP effectively processes sampled signals and accurately captures faults characteristics in power transformer diagnostics.

4.2. Feature Selection Algorithms

Feature selection refers to the process of identifying and removing irrelevant, less significant, and redundant features while selecting the most suitable inputs for a classification model.

- Particle Swarm Optimization (PSO) and Dwarf Mongoose Optimization (DMO) are employed for feature selection and hyperparameter tuning.
- The hybrid use of these two optimization methods ensures an efficient selection of the most relevant features while optimizing model parameters for an improved classification accuracy.

4.2.1. Particle Swarm Optimization (PSO)

The (PSO) algorithm [27], introduced by Eberhart and Kennedy, is a robust metaheuristic inspired by the collective behavior of social animals, such as birds in a flock or fish in a school. Each particle, representing a potential solution, updates its position by drawing

on its own experiences and the insights gained from its neighbors. This behavior reflects the way animals adapt their movements based on individual and collective knowledge. PSO is a versatile technique for solving a wide range of optimization problems, with each particle symbolizing a candidate solution. The process for identifying the optimal solution is outlined in a series of systematic steps [28]:

- (a) Initialization begins with creating a population of particles, each assigned a random velocity. The fitness of each particle is then evaluated to determine solution quality. The particle with the best performance is designated as the global best, while each particle retains its own best-known position as the local best.
- (b) During each iteration, particles move through the search space based on an updated velocity. This velocity is influenced by two main components: the global best position, representing the swarm's top-performing solution, and the particle's own local best position. Together, these guide the particle's trajectory toward more optimal regions of the solution space.
- (c) Position Update: After calculating the updated velocity, each particle's position is adjusted accordingly, allowing it to move through the search space. The updated velocity, which governs this movement, is determined using a specific mathematical formulation (referenced as Equation (15) [28]).

$$v_j^{i+1} = wv_j^{(i)} + (c_1 \times r_1 \times (\text{local best}_j - x_j^{(i)})) + (c_2 \times r_2 \times (\text{global best}_j - x_j^{(i)})) \quad v_{min} \leq v_j^{(i)} \leq v_{max} \quad (15)$$

- (d) The update on the global and local best values: If a particle achieves an improved performance based on the newly adjusted values, both its global and local best positions are updated. The process for determining and updating the local best for each particle follows the criteria outlined in Equation (16).

$$x_j^{i+1} = x_j^{(i)} + v_j^{(i+1)}; j = 1, 2, 3 \dots, n \quad (16)$$

- (e) Termination Check: If the specified stopping criteria are met, the current global best position is accepted as the final optimal solution. If not, the algorithm returns to the velocity update phase to proceed with further optimization.

4.2.2. The DMOA Model

The (DMOA) [29] is inspired by the adaptive behaviors of dwarf mongoose, incorporating constraints such as prey size, social hierarchy, and their seminomadic way of life. The social organization of dwarf mongooses is classified into three primary roles: alpha, scout, and babysitter groups. These roles collectively support migration and territorial exploration. The proposal DMOA methodology is structured into three key phases, as illustrated in Figure 6. The mathematical formulations used in the algorithm are adopted from Rf [30].

Population Initialization

As shown in Equation (17), the DMO algorithm starts by randomly generating an initial population of mongoose agents. Each agent's position is initialized within the problem's search space, bounded by predefined upper and lower limits.

$$X = \begin{bmatrix} x_{1,1}x_{1,2} \cdots x_{1,d-1}x_{1,d} \\ x_{2,1}x_{2,2} \cdots x_{2,d-1}x_{2,d} \\ \vdots x_{i,j} \vdots \\ x_{n,1}x_{n,2} \cdots x_{n,d-1}x_{n,d} \end{bmatrix} \quad (17)$$

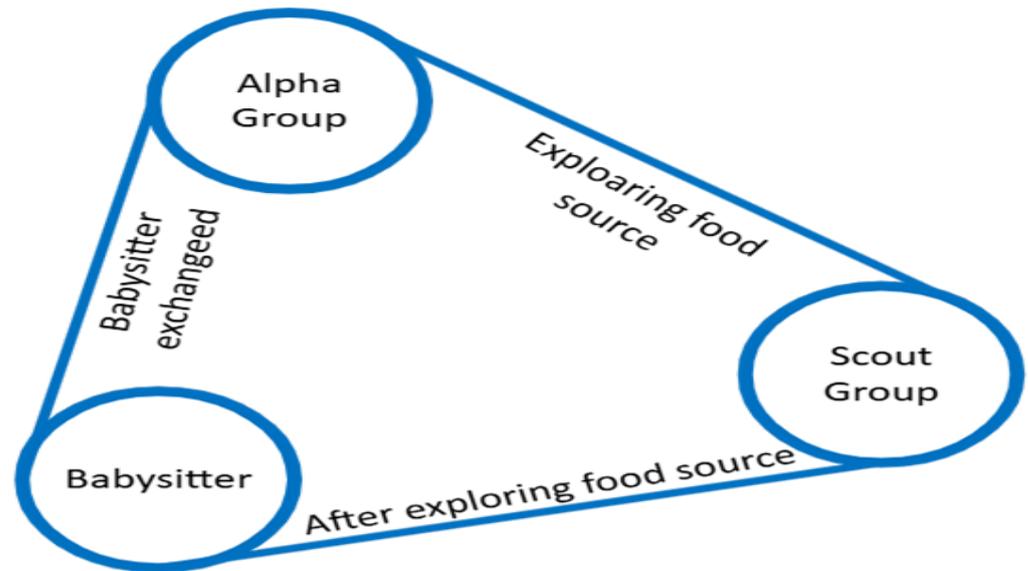


Figure 6. Optimization process steps of the specified DMOA.

The initial candidate population, represented as X , is generated randomly using Equation (18). Here, $x_{i,j}$ denotes the spatial position of the j th dimension within the i th population member. The parameter n specifies the population size, while d represents the dimensionality of the problem.

$$x_{i,j} = \text{unifrnd}(\text{Var}_{\text{Min}}, \text{Var}_{\text{Max}}, \text{Var}_{\text{Size}}) \quad (18)$$

The function *unifrnd* generates random numbers uniformly distributed between 0 and 1. Var_{Min} and Var_{Max} denote the lower and upper bounds of the problem, respectively. The variable Var_{Size} specifies the number of decision variables or the dimensions of the problem.

Alpha Group

Following initialization, each candidate solution undergoes a fitness evaluation. Equation (19) is used to calculate the probability corresponding to each solution's fitness. The alpha female (α) is then selected based on these computed probability values.

$$\alpha = \frac{\text{fit}_i}{\sum_{i=1}^n \text{fit}_i \text{fit}_i} \quad (19)$$

The size of the alpha group depends on the parameter $n - bs$. Additionally, bs represents the number of babysitters within the population. The vocalizations produced by the alpha female, referred to as *peep*, play a critical role in guiding the group while supporting internal harmony within the family unit.

Each mongoose initially rests in the first sleeping mound, represented by a value of 0. The DMOA employs Equation (20) to identify a potential food spot, simulating the process of exploration for sustenance.

$$X_{i+1} = X_i + \text{phi} * \text{peep} \quad (20)$$

Here, *phi* represents a random number uniformly distributed between -1 and 1 . The position of the sleeping mound is updated after each iteration, as described by Equation (21).

$$sm_i = \frac{\text{fit}_{i+1} - \text{fit}_i}{\max\{|\text{fit}_{i+1}|, |\text{fit}_i|\}} \quad (21)$$

Equation (22) calculates the mean value of the identified sleeping mounds, summarizing the average position based on the discoveries made during the search process.

$$\varphi = \frac{\sum_{i=1}^n sm_i}{n} \quad (22)$$

The variable n represents the number of babysitters within the population, reflecting their role in the social structure of the group.

Scout Group

Scouts are responsible for locating new resting mounds, as mongooses instinctively avoid reusing previous ones—encouraging continuous exploration. This scouting activity is carried out in parallel with foraging. The success of identifying a new mound is assessed based on exploration efficiency. As per the model, traveling a sufficient distance during foraging may lead to the discovery of an unvisited mound. The behavior of scout mongooses is mathematically represented by Equation (23).

$$X_{i+1} = \begin{cases} X_i - CF * phi * rand * \begin{bmatrix} X_i - \vec{M} \\ X_i - \vec{M} \end{bmatrix} & \text{if } \varphi_{i+1} > \varphi_i \\ X_i + CF * phi * rand * \begin{bmatrix} X_i - \vec{M} \\ X_i - \vec{M} \end{bmatrix} & \text{else} \end{cases} \quad (23)$$

In this context, $rand$ represents a randomly generated variable $[0, 1]$.

$CF = (1 - \frac{iter}{Max_iter})^{(2 \frac{iter}{Max_iter})}$ is utilized to control the collective–volitive movement of the mongoose group, gradually decreasing linearly as the number of iterations progress. The vector $\vec{M} = \sum_{i=1}^n \frac{X_i \times sm_i}{X_i}$ represents the directional movement of the mongoose as it navigates toward the newly discovered sleeping mound.

The Babysitters

The approach includes a rotating babysitting mechanism, enabling the alpha female to guide the group during daily foraging. The count of babysitters is derived from the total population and influences the optimization process by effectively reducing the active number of individuals. This behavior is mimicked in the algorithm by scaling down the effective population size according to the proportion assigned to babysitting duties.

During the babysitter rotation, data related to scouting and food discovery are refreshed, and babysitters are given a fitness score of zero. As iterations proceed, the average fitness of the alpha group gradually declines, restricting their collective movement and encouraging a stronger exploitation in the search process.

4.3. Classification Techniques

Three machine learning models—(DT), and (RF), and (SVM)—were selected for this study because of their proven capabilities in addressing the challenges of transformer fault detection. These models excel in managing nonlinear relationships, handling imbalanced fault class distributions, and meeting the demands of real-time classification.

- The optimized feature sets are classified using (DT), (RF), and (SVM).
- A 5-fold cross-validation strategy is used to validate performance, ensuring robust and generalizable results.

4.3.1. Decision Tree

Decision Trees are particularly effective for detecting transformer faults, thanks to their capability to manage nonlinear patterns and unbalanced class distributions. They operate by recursively partitioning the dataset based on feature thresholds, aiming to

produce purer subsets at each terminal node. This results in a tree-like structure that is both interpretable and informative, showing clear associations between voltage-based features and fault types—helping to identify potential causes of transformer issues [31].

These models are also known for their computational efficiency, making them suitable for real-time applications. Additionally, their inherent flexibility allows them to mitigate overfitting, especially when dealing with complex or noisy data. In this study, the tree's maximum depth is fine-tuned using 5-fold cross-validation to maintain a strong balance between performance and generalization.

Mathematical Representation [32,33]:

A decision tree classifies input by systematically splitting the feature space. At each decision node, the algorithm selects the feature and corresponding threshold t that yields the most effective division into two subsets, typically minimizing impurity measures such as the Gini index:

$$G = 1 - \sum_{k=1}^K p_k^2 \quad (24)$$

Here, p_k represents the relative frequency of class k instances at a given node. The splitting process proceeds iteratively until a stopping condition, such as reaching a maximum tree depth or a minimum number of samples per leaf, is satisfied.

4.3.2. Support Vector Machine (SVM)

Support Vector Machines (SVMs) are powerful supervised learning algorithms well suited for high-dimensional and complex classification tasks, making them highly applicable for diagnosing transformer faults. They operate by determining a separating hyperplane that maximizes the margin between classes, where support vectors—the closest data points to the boundary—play a central role in defining that margin. This margin-based approach improves the model's robustness to noise and variability in voltage signals, which is essential for real-world fault detection scenarios [34,35].

SVMs also perform well in handling class overlap, a common challenge in transformer fault datasets. In this work, we utilize an SVM with a radial basis function (RBF) kernel to capture the nonlinear relationships between voltage features and fault types. To enhance model accuracy, the primary RBF hyperparameters— C (the regularization factor) and γ (which controls the kernel's influence)—are tuned using 5-fold cross-validation.

Mathematical Formulation [34,36]:

The SVM classification task involves solving an optimization problem aimed at identifying the hyperplane that best separates different fault categories in the feature space.

$$\min_{w,b} \frac{1}{2} |w|^2 \text{ subject to } y_i (w^\top x_i + b) \geq 1, \forall i \quad (25)$$

Here, w represents the weight vector, b is the bias term, y_i denotes the class labels, and x_i are the feature vectors. The kernel trick transforms input data for nonlinear classification into a higher-dimensional space. The radial basis function (RBF) kernel is expressed as follows:

$$K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \quad (26)$$

where γ controls the width of the kernel.

4.4. Random Forest (RF)

Random Forest (RF) is an ensemble-based classification technique that aggregates the predictions of numerous decision trees to improve the overall model accuracy and

robustness. Each tree is trained on random subsets of the data, with final predictions based on using a majority rule to determine class labels and the mean prediction for continuous outputs [37].

Key hyperparameters include **NumTrees** (number of trees), **MinLeafSize** (minimum samples per leaf node to prevent overfitting), and **MaxNumSplits** (maximum allowable splits per tree to control depth). These parameters were optimized using 5-fold cross-validation for this study. RF's ability to manage nonlinear relationships, imbalanced datasets, and high-dimensional data makes it suitable for transformer fault classification.

In the Mathematical Model, for classification, the final class prediction is [38,39] as follows:

$$\hat{y} = \text{mode}(y_1, y_2, \dots, y_T) \quad (27)$$

where \hat{y} is a final predicted class from the t-th tree, and T: the total number of trees (NumTrees).

$$\hat{y} = \frac{1}{T} \sum t_1^T y_t \quad (28)$$

For regression, it is as follows:

Node splits minimize impurity using the Gini index:

$$G = 1 - \sum i_1 k p^2 \quad (29)$$

where G is the Gini index, and k is the number of classes, the proportion of samples belonging to class i . Figure 7 illustrates the schematic diagram of the proposed technique.

4.5. Performance Evaluation Metrics

To assess the performance of the machine learning models, a broad range of evaluation metrics was employed. These metrics provide meaningful insights into model accuracy and reliability, supporting an in-depth comparison to determine the most effective approach. The evaluation framework is based on the confusion matrix, which compares predicted outcomes with the actual class labels.

In binary classification settings, the confusion matrix comprises four key components:

- True Positives (TP): Faulty cases correctly predicted as faulty.
- True Negatives (TN): Healthy cases correctly predicted as healthy.
- False Positives (FP): Healthy cases incorrectly predicted as faulty (Type I error).
- False Negatives (FN): Faulty cases incorrectly predicted as healthy (Type II error).

Table 5 summarizes the performance outcomes of the three machine learning models, assessed using a 5-fold cross-validation approach.

Table 5. A performance evaluation of three machine learning models using 5-fold cross-validation.

Metric	Formula
Sensitivity (Recall)	$Sensitivity = \frac{TP}{TP+FN}$
Specificity	$Specificity = \frac{TN}{TN+FP}$
Precision	$Precision = \frac{TP}{TP+FP}$
F1 Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
Negative Predictive Value (NPV)	$NPV = \frac{TN}{TN+FN}$

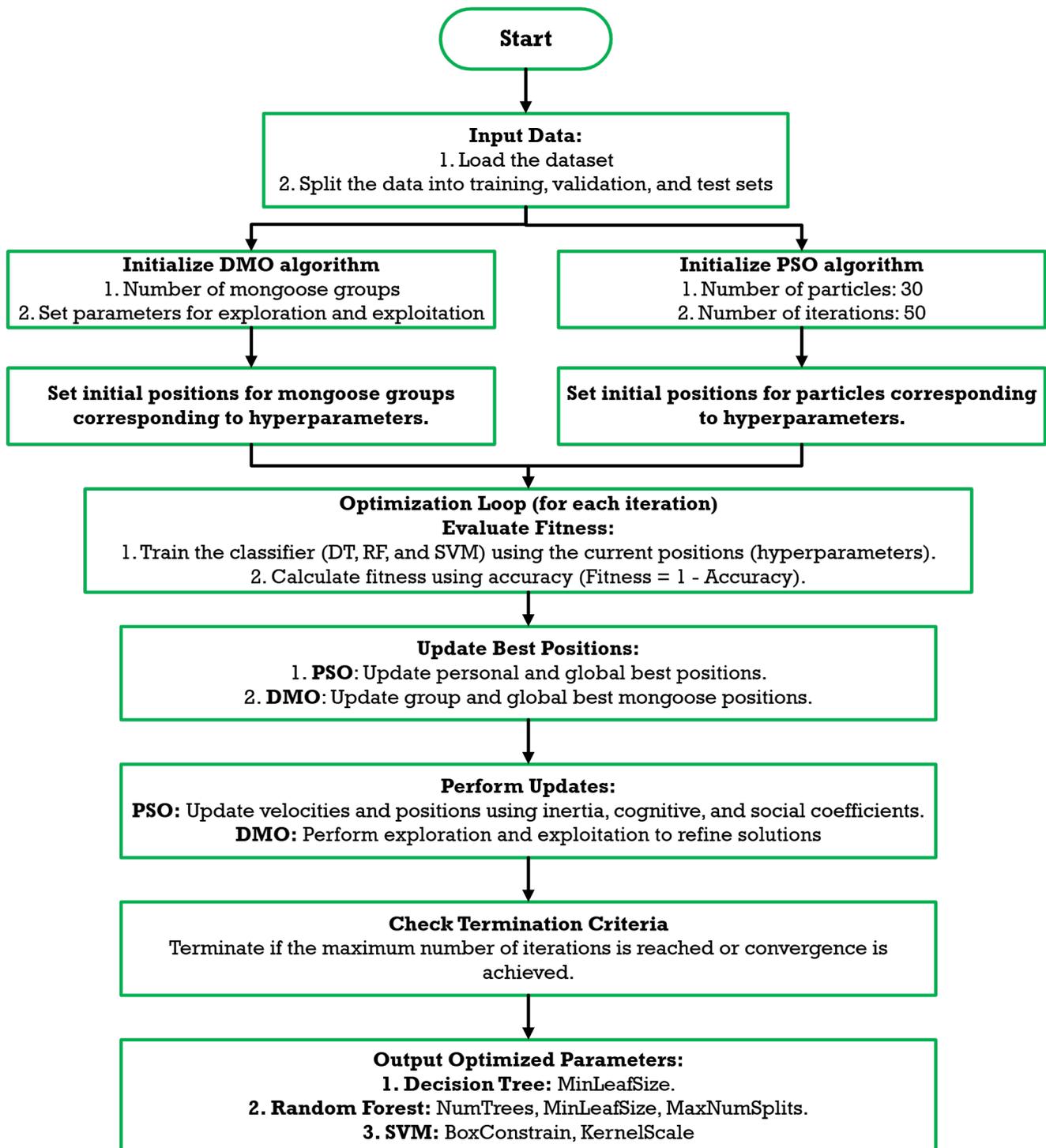


Figure 7. A schematic diagram of the proposal model.

These evaluation metrics provide a well-rounded view of each model's effectiveness, supporting a thorough and balanced assessment of classification performance.

- Sensitivity (Recall): Measures how effectively the model identifies faulty transformers.
- Specificity: Indicates the model's accuracy in correctly classifying healthy transformers.
- Precision: Denotes the proportion of true fault detections among all instances predicted as faulty.
- Negative Predictive Value (NPV): Represents the share of truly healthy cases among those predicted as non-faulty.

- Accuracy: Reflects the overall proportion of correctly classified samples, including both healthy and faulty categories.
- F1 Score: Calculates the harmonic mean of precision and recall, offering a balanced metric, which is especially useful for datasets with a class imbalance.

By evaluating this comprehensive set of metrics, the authors strive to identify the model that achieves an optimal balance between sensitivity, specificity, and accuracy while maximizing the F1 score. This approach ensures the development of a reliable and effective fault detection system.

The proposed fault detection and classification methodology for power transformers, illustrated in Figure 8, follows a structured workflow. It begins with current signal acquisition, followed by feature extraction using Orthogonal Matching Pursuit (OMP) and Discrete Wavelet Transform (DWT) to capture key time–frequency characteristics. DMO and PSO then optimize feature selection, reducing dimensionality and improving classifier performance.

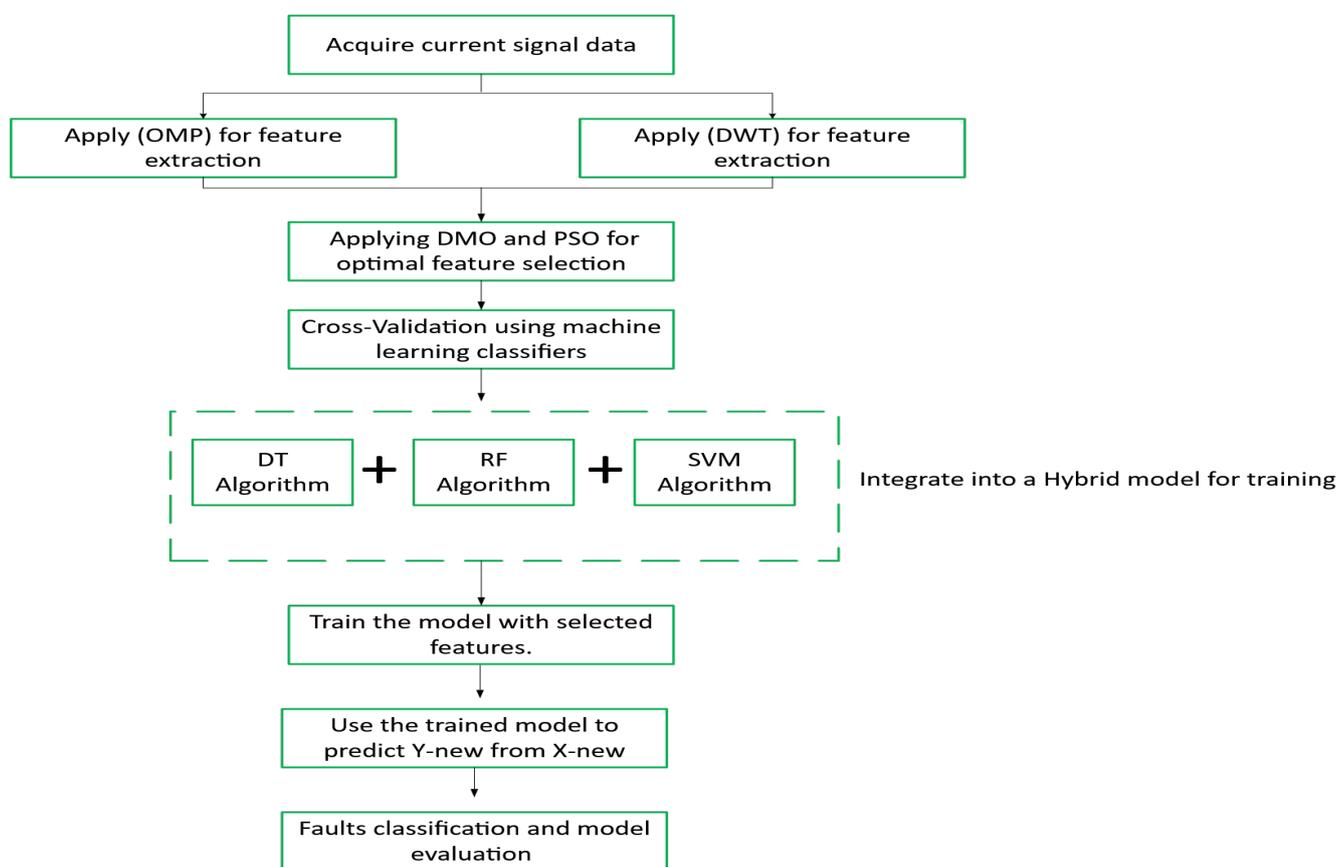


Figure 8. A schematic diagram of the proposal methodology.

A hybrid machine learning model combining (RF), (DT), and (SVM) is trained on the selected features to enhance classification accuracy. The trained model predicts fault conditions by mapping unseen input features (X_{new}) to fault categories (Y_{new}). Finally, model performance is evaluated using Accuracy, Precision, Recall, F1 Score, Specificity, and the Negative Predictive Value (NPV), ensuring a reliable fault diagnosis system.

5. Results and Discussion

5.1. Performance Evaluation

The performance of different classifiers optimized with PSO and DMO was evaluated using Accuracy, Precision, Recall, Sensitivity, F1 Score, Specificity and NPV. The results

demonstrate that DMO-optimized models consistently outperform PSO-optimized models. Figure 9, Tables 6 and 7 summarize the classification performance of different approaches.

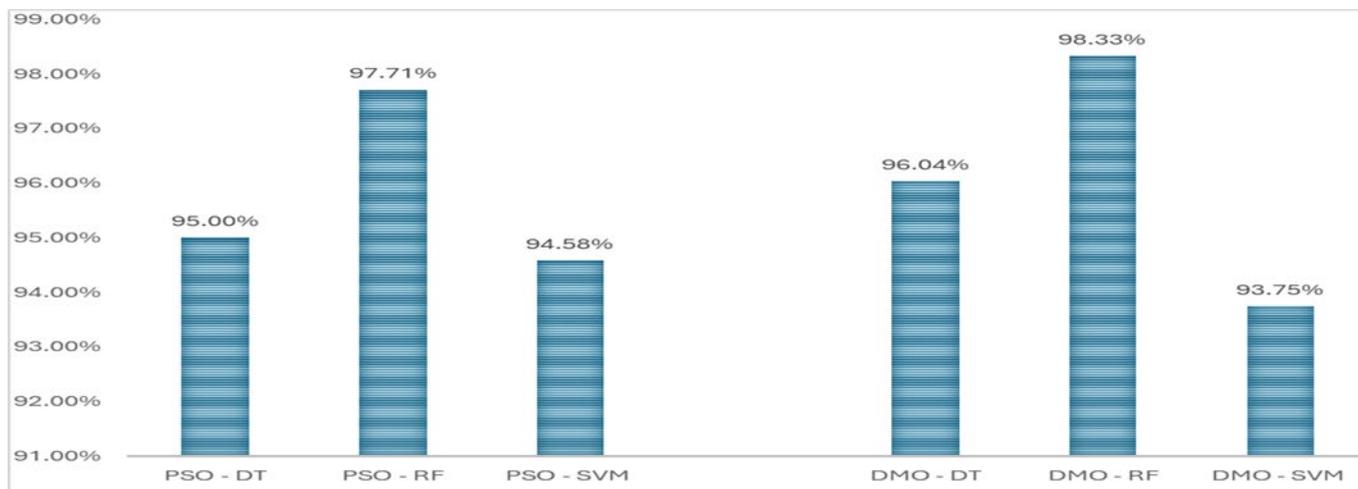


Figure 9. A comparative performance analysis of the proposal methods.

Table 6. The performance of PSO-based models.

Model	Accuracy	Precision	Sensitivity (Recall)	F1 Score	Specificity	Negative Predictive Value (NPV)
PSO-DT	95.00%	97.72%	96.25%	96.98%	88.75%	82.56%
PSO-RF	97.71%	98.02%	99.25%	98.63%	90.00%	96.00%
PSO-SVM	94.58%	95.61%	98.00%	96.79%	77.50%	88.57%

Table 7. Performance Metrics of PSO-Based Models.

Model	Accuracy	Precision	Sensitivity (Recall)	F1 Score	Specificity	Negative Predictive Value (NPV)
DMO-DT	96.04%	95.90%	99.50%	97.67%	78.75%	96.92%
DMO-RF	98.33%	98.80%	99.28%	99.04%	92.31%	95.24%
DMO-SVM	93.75%	95.57%	97.00%	96.28%	77.50%	83.78%

A paired *t*-test was conducted to compare the PSO- and DMO-based models. The resulting *p*-values confirmed that the performance differences are statistically significant, supporting the reliability of the observed improvements.

The confusion matrix provides insight into the model's prediction accuracy for each class. Figures 10 and 11 illustrate the confusion matrix for the best-performing model (DMO-RF), where, as follows:

True Positives (TP) and True Negatives (TN) dominate, indicating a high classification accuracy.

False Positives (FP) and False Negatives (FN) are minimal, suggesting the model's robustness in distinguishing between faulty and healthy conditions.

The misclassification rate is significantly lower for DMO-optimized models, reinforcing the effectiveness of the proposed approach.

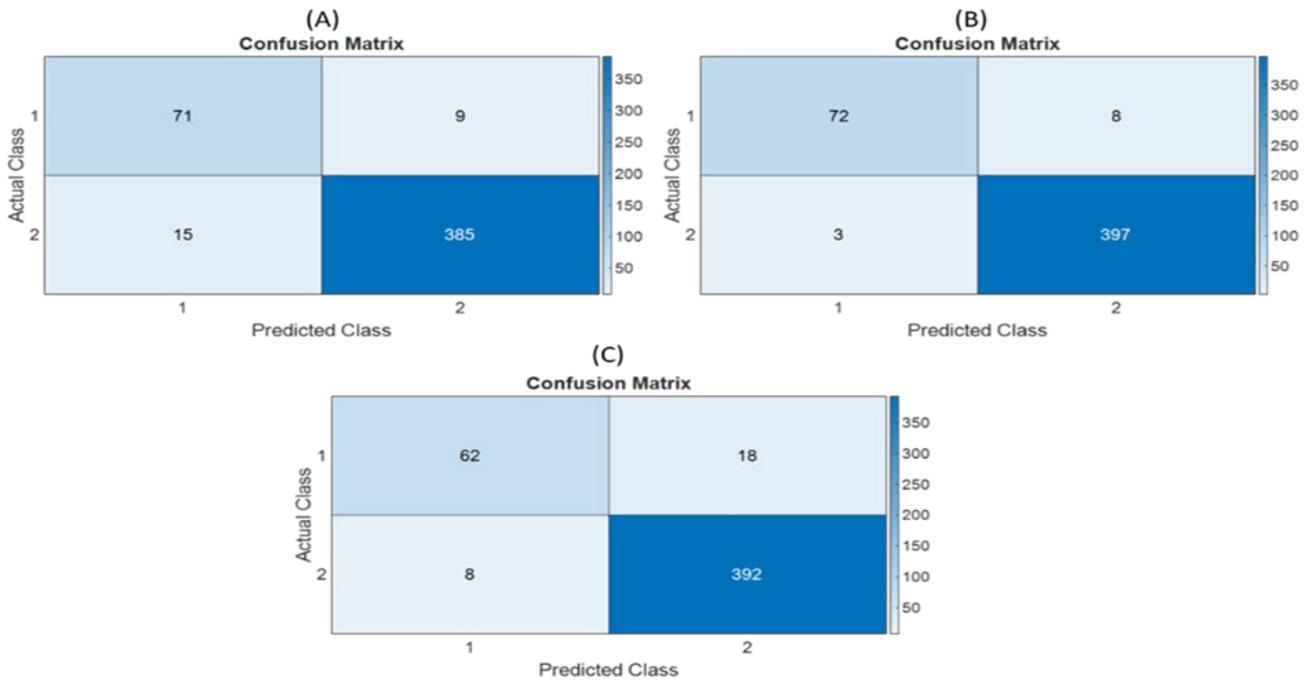


Figure 10. Confusion matrices for PSO-based models. (A) PSO-DT, (B) PSO-RF, and (C) PSO-SVM.

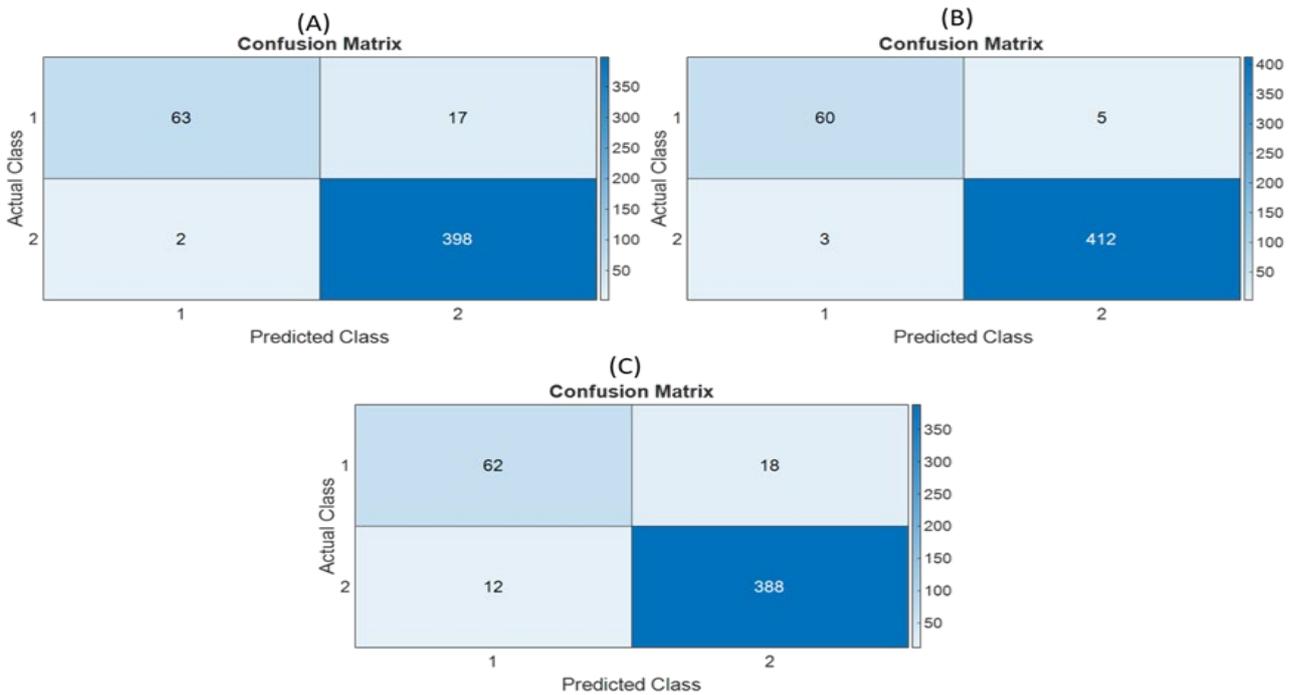


Figure 11. Confusion matrices for DMO-based models. (A) DMO-DT, (B) DMO-RF, and (C) DMO-SVM.

Figure 12 shows the ROC curves for models optimized using PSO and DMO, highlighting the balance between the True Positive Rate (TPR) and False Positive Rate (FPR). The RF model performs best with an AUC of 0.99, followed by SVM (0.97) and DT (0.96). Both optimization techniques yield similar results, demonstrating their effectiveness in enhancing classification accuracy. The minimal AUC variations confirm the models' robustness and reliability in distinguishing between classes.

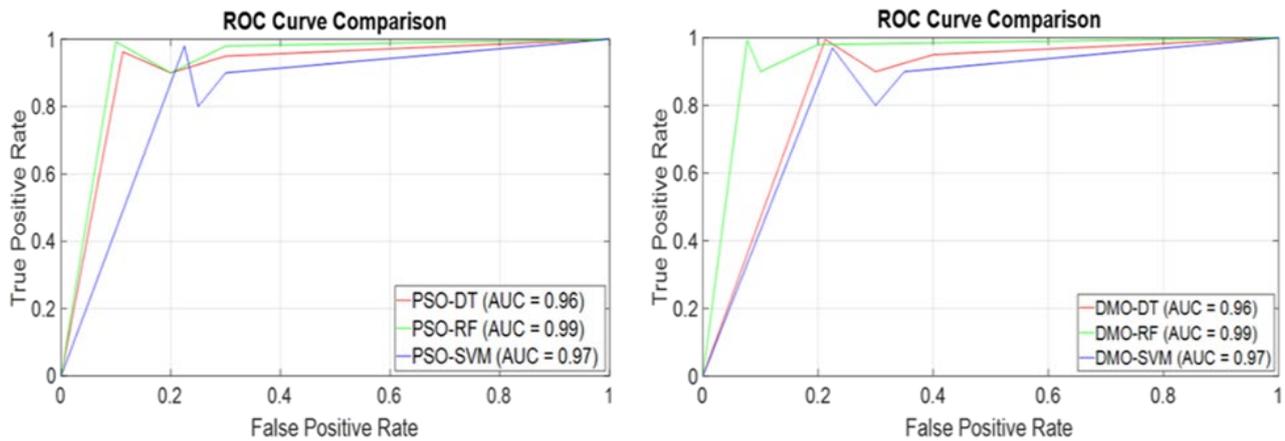


Figure 12. A combined ROC curve comparison for PSO and DMO classifiers.

5.2. Comparative Analysis of Optimization Techniques

The effectiveness of the PSO and DMO optimization techniques was assessed by comparing their impact on model performance, convergence speed, and computational efficiency. PSO, while widely used, tends to get trapped in local optima and exhibits a slower convergence in high-dimensional feature spaces. DMO, on the other hand, introduces adaptive strategies that improve the balance between exploration and exploitation, allowing for a more optimal feature selection and a faster convergence.

Table 8 summarizes the comparative performance of PSO and DMO in terms of accuracy, convergence time, and iterations required to achieve optimal solutions.

Table 8. The comparative performance of PSO and DMO.

Optimization Method	Best Accuracy (%)	Iterations to Optimal Solution
PSO	97.71	85
DMO	98.33	60

As shown in Table 8, DMO consistently outperforms PSO in both accuracy and computational efficiency. The lower number of required iterations indicates a more effective search process, making it a preferable choice for real-time transformer fault detection.

5.3. Practical Implications

The proposed DMO-optimized framework is highly applicable for industrial use, offering a scalable and real-time solution for power transformer fault detection. Key considerations for practical deployment include the following:

- **An Integration with SCADA Systems:** The model can be deployed in Supervisory Control and Data Acquisition (SCADA) systems to enable automated, real-time fault monitoring.
- **The Handling of Sensor Constraints:** Industrial transformers are exposed to sensor noise and environmental disturbances; robust pre-processing techniques (e.g., wavelet denoising) can enhance signal reliability.
- **An Adaptability to Different Transformer Types:** The model's adaptive feature selection ensures that it generalizes well across different transformer sizes and configurations.
- **The Reducing of False Alarms:** By incorporating advanced metaheuristic optimization, the framework minimizes misclassification rates, ensuring a higher protection reliability.

5.4. Real-World Applicability and Industrial Integration

This framework is designed to support integration into real-world power systems. Key practical challenges and considerations include the following:

- **Sensor Noise and Environmental Interference:** Addressed using pre-processing and denoising techniques.
- **Varying Load Conditions:** The model is trained on diverse operating scenarios to improve generalization.
- **Integration with Protection Systems:** The classifier output can support real-time decision-making in SCADA environments.
- **Latency and Computational Efficiency:** DMO enables a faster convergence, supporting real-time deployment.
- **Adaptability:** The proposed framework can be adjusted for different transformer ratings and topologies.

6. Conclusions

This study presents a hybrid optimization-based machine learning framework for power transformer fault detection. The DMO-optimized RF model achieved a 98.33% accuracy, outperforming other classifiers. The key contributions include the following:

- An effective hybrid metaheuristic approach integrating PSO and DMO for optimal feature selection.
- A robust experimental validation setup, ensuring a realistic fault simulation.
- A better computational efficiency than deep learning-based methods, making it practical for industrial deployment.

Limitations and Future Work

Despite its strong performance, some limitations exist, as follows:

- The scalability for high-voltage transformers remains to be explored.
- Real-time implementation using embedded hardware (FPGA and IoT) should be tested.
- A hybridization with deep learning techniques could be considered for further improvements.

This study bridges the gap between metaheuristic optimization and real-world transformer fault detection, providing an effective solution for modern power grid applications.

Author Contributions: Conceptualization, M.A., M.S. and F.A.; methodology, M.A.; software, M.S.; validation, M.A., M.P., F.A. and M.S.; formal analysis, M.A., M.P., F.A. and M.S.; investigation, M.A., M.P., F.A. and M.S.; resources, M.A., M.P., F.A. and M.S.; data curation, M.A., M.P., F.A. and M.S.; writing—original draft preparation, M.A., M.P., F.A. and M.S.; writing—review and editing, M.A., M.P., F.A. and M.S.; visualization, M.A., M.P., F.A. and M.S.; supervision, F.A.; project administration, M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research forms a part of the Ph.D. work of the corresponding author, Mohammed Alenezi.

Data Availability Statement: The dataset supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the support of Cardiff University, School of Engineering, for covering the Article Processing Charge (APC) associated with this publication.

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

References

1. Alenezi, M.; Anayi, F.; Packianather, M.; Shouran, M. Enhancing Transformer Protection: A Machine Learning Framework for Early Fault Detection. *Sustainability* **2024**, *16*, 10759. [\[CrossRef\]](#)
2. Maseko, N.S.; Thango, B.A.; Mabunda, N. Fault Detection in Power Transformers Using Frequency Response Analysis and Machine Learning Models. *Appl. Sci.* **2025**, *15*, 2406. [\[CrossRef\]](#)
3. Abbasi, A.R. Fault detection and diagnosis in power transformers: A comprehensive review and classification of publications and methods. *Electr. Power Syst. Res.* **2022**, *209*, 107990. [\[CrossRef\]](#)
4. Ahmadi, H.; Vahidi, B.; Nematollahi, A.F. A simple method to detect internal and external short-circuit faults, classify and locate different internal faults in transformers. *Electr. Eng.* **2021**, *103*, 825–836. [\[CrossRef\]](#)
5. Et, V.K.S. A Review of Various Protection Schemes Of Power Transformers. *Turk. J. Comput. Math. Educ. (TURCOMAT)* **2021**, *12*, 3220–3228. [\[CrossRef\]](#)
6. Ekojono; Prasojo, R.A.; Apriyani, M.E.; Rahmanto, A.N. Investigation on machine learning algorithms to support transformer dissolved gas analysis fault identification. *Electr. Eng.* **2022**, *104*, 3037–3047. [\[CrossRef\]](#)
7. Guezzaz, A.; Asimi, Y.; Azrou, M.; Asimi, A. Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection. *Big Data Min. Anal.* **2021**, *4*, 18–24. [\[CrossRef\]](#)
8. Charbuty, B.; Abdulazeez, A. Classification Based on Decision Tree Algorithm for Machine Learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [\[CrossRef\]](#)
9. Khajavi, H.; Rastgoo, A. Predicting the carbon dioxide emission caused by road transport using a Random Forest (RF) model combined by Meta-Heuristic Algorithms. *Sustain. Cities Soc.* **2023**, *93*, 104503. [\[CrossRef\]](#)
10. Dabiri, H.; Farhangi, V.; Moradi, M.J.; Zadehmohamad, M.; Karakouzian, M. Applications of Decision Tree and Random Forest as Tree-Based Machine Learning Techniques for Analyzing the Ultimate Strain of Spliced and Non-Spliced Reinforcement Bars. *Appl. Sci.* **2022**, *12*, 4851. [\[CrossRef\]](#)
11. Ghosh, A.; Maiti, R. Soil erosion susceptibility assessment using logistic regression, decision tree and random forest: Study on the Mayurakshi river basin of Eastern India. *Environ. Earth Sci.* **2021**, *80*, 328. [\[CrossRef\]](#)
12. Alhijawi, B.; Awajan, A. *Genetic Algorithms: Theory, Genetic Operators, Solutions, and Applications*; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2024. [\[CrossRef\]](#)
13. Liu, Y.; As'arry, A.; Hassan, M.K.; Hairuddin, A.A.; Mohamad, H. Review of the grey wolf optimization algorithm: Variants and applications. *Neural Comput. Appl.* **2024**, *36*, 2713–2735. [\[CrossRef\]](#)
14. Ahmad, M.F.; Isa, N.A.M.; Lim, W.H.; Ang, K.M. Differential evolution: A recent review based on state-of-the-art works. *Alex. Eng. J.* **2022**, *61*, 3831–3872. [\[CrossRef\]](#)
15. Indrawati, A.; Wahyuni, I.N. Enhancing Machine Learning Models through Hyperparameter Optimization with Particle Swarm Optimization. In Proceedings of the 2023 10th International Conference on Computer, Control, Informatics and its Applications: Exploring the Power of Data: Leveraging Information to Drive Digital Innovation, IC3INA 2023, Bandung, Indonesia, 4–5 October 2023; pp. 244–249. [\[CrossRef\]](#)
16. Mohamed, N.; Almutairi, R.L.; Abdelrahim, S.; Alharbi, R.; Alhomayani, F.M.; Elamin Elnaim, B.M.; Elhag, A.A.; Dhakal, R. Automated Laryngeal Cancer Detection and Classification Using Dwarf Mongoose Optimization Algorithm with Deep Learning. *Cancers* **2024**, *16*, 181. [\[CrossRef\]](#)
17. Thomas, J.B.; Chaudhari, S.G.; Shihabudheen, K.V.; Verma, N.K. CNN-Based Transformer Model for Fault Detection in Power System Networks. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3238059. [\[CrossRef\]](#)
18. Lu, S.; Gao, W.; Hong, C.; Sun, Y. A newly-designed fault diagnostic method for transformers via improved empirical wavelet transform and kernel extreme learning machine. *Adv. Eng. Inform.* **2021**, *49*, 101320. [\[CrossRef\]](#)
19. Cao, H.; Zhou, C.; Meng, Y.; Shen, J.; Xie, X. Advancement in transformer fault diagnosis technology. *Front. Energy Res.* **2024**, *12*, 1437614. [\[CrossRef\]](#)
20. Shang, H.; Liu, Z.; Wei, Y.; Zhang, S. A Novel Fault Diagnosis Method for a Power Transformer Based on Multi-Scale Approximate Entropy and Optimized Convolutional Networks. *Entropy* **2024**, *26*, 186. [\[CrossRef\]](#)
21. Li, Q.; Luo, H.; Cheng, H.; Deng, Y.; Sun, W.; Li, W.; Liu, Z. Incipient Fault Detection in Power Distribution System: A Time-Frequency Embedded Deep Learning Based Approach. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2507914. [\[CrossRef\]](#)
22. Cui, M.; Wang, Y. Fault detection of rotating machinery based on wavelet transform and improved deep neural network. In Proceedings of the 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), Liuzhou, China, 20–22 November 2020; pp. 449–454. [\[CrossRef\]](#)
23. Li, G.; Wei, M.; Shao, H.; Liang, P.; Duan, C. Wavelet Knowledge-Driven Transformer for Intelligent Machinery Fault Detection with Zero-Fault Samples. *IEEE Sens. J.* **2024**, *24*, 35986–35996. [\[CrossRef\]](#)
24. Ibrahim, A.; Anayi, F.; Packianather, M.; Alomari, O.A. New Hybrid Invasive Weed Optimization and Machine Learning Approach for Fault Detection. *Energies* **2022**, *15*, 1488. [\[CrossRef\]](#)

25. Pati, Y.C.; Rezaifar, R.; Krishnaprasad, P.S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; Volume 1, pp. 40–44. [[CrossRef](#)]
26. Huang, L.; Asteris, P.G.; Koopialipour, M.; Armaghani, D.J.; Tahir, M.M. Invasive weed optimization technique-based ANN to the prediction of rock tensile strength. *Appl. Sci.* **2019**, *9*, 5372. [[CrossRef](#)]
27. Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. In Proceedings of the MHS'95, Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 4–6 October 1995; pp. 39–43. [[CrossRef](#)]
28. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. Proceedings of ICNN'95—International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995. [[CrossRef](#)]
29. Al Dawsari, S.; Anayi, F.; Packianather, M. Techno-economic analysis of hybrid renewable energy systems for cost reduction and reliability improvement using dwarf mongoose optimization algorithm. *Energy* **2024**, *313*, 133653. [[CrossRef](#)]
30. Agushaka, J.O.; Ezugwu, A.E.; Abualigah, L. Dwarf Mongoose Optimization Algorithm. *Comput. Methods Appl. Mech. Eng.* **2022**, *391*, 114570. [[CrossRef](#)]
31. Mitu, M.M.; Arefin, S.; Saurav, Z.; Hasan, M.A.; Farid, D.M. Pruning-Based Ensemble Tree for Multi-Class Classification. In Proceedings of the 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024, Dhaka, Bangladesh, 2–4 May 2024; pp. 481–486. [[CrossRef](#)]
32. Quinlan, J.R. Induction of Decision Trees. *Mach Learn* **1986**, *1*, 81–106. [[CrossRef](#)]
33. Kotsiantis, S.B. Decision trees: A recent overview. *Artif. Intell. Rev.* **2013**, *39*, 261–283. [[CrossRef](#)]
34. Cortes, C.; Vapnik, V.; Saitta, L. *Support-Vector Networks Editor*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1995.
35. Iqbal, G.M.; Rosenberger, J.; Rosenberger, M.; Ha, L.; Anoruo, E.; Gregory, S.; Mazzone, T. A Support Vector Machine based Logistic Regression Model Approach for Classification Problems. In Proceedings of the IISE Annual Conference and Expo 2023, New Orleans, LA, USA, 21–23 May 2023. [[CrossRef](#)]
36. Chauhan, V.K.; Dahiya, K.; Sharma, A. Problem formulations and solvers in linear SVM: A review. *Artif. Intell. Rev.* **2019**, *52*, 803–855. [[CrossRef](#)]
37. Chen, X.; Ji, N.; Qin, X.; Zhang, M.; Chen, X.; Jiang, C.; Tao, K. Transformer Fault Diagnosis Based on the Improved Sparrow Search Algorithm and Random Forest Feature Selection. In Proceedings of the 2024 3rd International Conference on Energy and Electrical Power Systems, ICEEPS 2024, Guangzhou, China, 14–16 July 2024; pp. 1086–1091. [[CrossRef](#)]
38. Khatib, T.; Arar, G. Identification of Power Transformer Currents by Using Random Forest and Boosting Techniques. *Math. Probl. Eng.* **2020**, *2020*, 1269367. [[CrossRef](#)]
39. Shah, A.M.; Bhalja, B.R. Fault discrimination scheme for power transformer using random forest technique. *IET Gener. Transm. Distrib.* **2016**, *10*, 1431–1439. [[CrossRef](#)]

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