Journal port Science Research
https://jpsr.uruk.edu.iq/index.php/jport
Volume 9, issue 2. 2026

Benchmarking Binary Metaheuristics for Joint Feature Selection and SVM Tuning in Biomedical Classification

Ahmed Majid Taha [a&b]

[a] College of Biomedical Informatics, University of Information Technology and Communications, Baghdad, Iraq
[b] Soft Computing and Data Mining Center, Universiti Tun Hussein Onn Malaysia 86400 Parit Raja, Batu Pahat Johor, Malaysia

ARTICLE INFO

Keywords:

Biomedical classification; Joint optimization; Feature selection; Hyperparameter tuning; Binary metaheuristics; Support .vector machine

ABSTRACT

Hyperparameter optimization and feature selection are closely related aspects of biomedical classification systems, especially when samples are high-dimensional and limited in sample size. Traditional pipelines normally tackle these components separately, though they interact heavily in kernel-based classifiers like Support Vector Machines (SVMs). This paper presents a unified optimization model that concurrently selects informative biomedical features and optimizes SVM hyperparameters, and compares three high-level binary metaheuristic algorithms (Binary Harris Hawks Optimization (BHHO), Binary Differential Evolution (BDE), and Binary Marine Predators Algorithm (BMPA)) in the same set of experimental conditions. Both candidate solutions combine a binary feature-selection vector with continuous SVM parameters to allow exploration of the mixed search space. The proposed framework is tested against various biomedical data to perform different diagnostic tasks. The evaluation is performed experimentally in terms of AUC, F1-score, accuracy, and the feature reduction rate. Results demonstrate consistent improvements over baseline methods, with the highest overall discriminative performance recorded with BHHO, while BMPA produces the most compact feature subsets. This result shows that combined optimization using the proposed binary metaheuristics provides a powerful and interpretable method of diagnosing biomedical problems.

1. INTRODUCTION

Machine learning (ML) systems are now a significant part of the modern biomedical diagnosis system, where they can assist in identifying disease, risk stratification, and making clinical decisions. However effective, biomedical data is frequently a problem because of its high-dimensionality, redundancy or noise, disproportionate classes, and the number of labeled specimens. They tend to deteriorate model generalization and make it more difficult to interpret, which is especially important in the medical domain. To overcome these difficulties, feature selection (FS), in many cases, is used to extract a set of informative features that have the greatest contribution to the predictive task. With effective feature selection, biomedical scenarios can improve classification precision, decrease overfitting, and promote clinical interpretability, with a focus on the identification of the relevant biomarkers (Gao et al., 2023; Khan et al., 2024). Wrapper-based FS considers subsets of candidate features and optimizes them using a learning algorithm; they are particularly appealing because of their high performance, though they add to the complexity of combinatorial optimization. The use of Support Vector machines (SVMs) is a popular approach in biomedical classification due to their good theoretical basis and the ability to work well with small and medium-sized datasets. SVM is, however, very sensitive to the selection of kernel hyperparameters, especially with the nonlinear kernels like Radial Basis Function (RBF). Poorly chosen hyperparameter values may severely reduce the diagnostic accuracy, despite the choice of informative features (Vincent et al., 2023).

E-mail address:
dr.ahmed_majid@uoitc.edu.iq

Received 30th December 2025، Accepted 1nd January 2026

10.36371/port.2026.2.3

The majority of available research focuses on feature selection and hyperparameter optimization as independent preprocessing tasks; this division does not take into consideration the high dependency between the chosen feature subset and the best hyperparameter setting. Practically, a set of optimal features with a particular set of hyperparameters can be non-optimal with a different set of hyperparameters. Subsequently, solutions that have been independently optimized might not yield global optimality. As a solution to this shortcoming, recent studies have started to consider joint optimization models, where feature selection and hyperparameter optimization are considered one mixed-variable optimization problem (Akyol et al., 2024; Yokoyama et al., 2025). Such joint formulations are especially suitable for the metaheuristic optimization algorithms because of their flexibility, gradient-free character, and avoidance of local optima. Metaheuristics have binary variants that allow them to model feature inclusion and exclusion directly, with continuous elements allowed to optimize hyperparameters of the classifier at the same time. Newer metaheuristics have shown better exploration-exploitation ratio and convergence in difficult search spaces despite the widespread use of classical methods like Particle Swarm Optimization and Genetic Algorithms. This study focuses on three modern and competitive algorithms: Binary Harris Hawks Optimization (BHHO), Binary Differential Evolution (BDE), and Binary Marine Predators Algorithm (BMPA). Although the results of these algorithms have been promising on a case-by-case basis, their relative efficacy with respect to combined feature selection and SVM hyperparameter optimization when it comes to biomedical diagnosis has not been thoroughly explored.

Contributions

This study made the following contributions to knowledge:

a. Formulation of the feature selection and SVM hyperparameter tuning tasks as a single mixed binary-continuous optimization problem, explicitly modeling their interdependence.
b. Adaptation and evaluation of three recent and competitive algorithms (BHHO, BDE, and BMPA) under a unified experimental protocol for biomedical classification.
c. Comparison of the evaluated algorithms using identical computational budgets, cross-validation protocols, and non-parametric statistical tests to ensure methodological rigor.
d. Evaluation of the predictive performance, feature compactness, convergence behavior, and statistical significance of the evaluated models across multiple biomedical datasets.
e. Provision of actionable insights regarding algorithm selection based on dataset characteristics and interpretability requirements.

This paper is not an attempt to come up with a hybrid metaheuristic algorithm, as such; rather, it is a controlled comparative benchmarking of three current binary optimizers as part of a single joint optimization setup. The rest of this paper is structured in the following way: Section 2 examines the recent progress in the biomedical feature selection and binary metaheuristic optimization. Section 3 presents the problem of joint feature selection and hyperparameter optimization as a mixed-variable optimization problem. Section 4 outlines the proposed binary modifications of BHHO, BDE, and BMPA in the single optimization model. Section 5 describes the experimental design, such as datasets, measures of evaluation, and statistical verification processes. Section 6 presents the comparison of the empirical findings. Section 7 addresses possible threats to validity; Section 8 concludes the paper study.

2. Related Work

2.1 Feature selection in biomedical machine learning

Biomedical datasets are usually characterized by a high level of dimensionality, redundant variables, and a small sample size, which adversely impact the classification process and model comprehension. The task of feature selection (FS) is consequently considered a key step in preprocessing biomedical machine learning, as the purpose is to extract the most informative set of features with the lowest possible noise and lowest computational cost. Recent findings suggest that successful FS has the potential to enhance the accuracy and strength of diagnostic outcomes, especially when dealing with disease prediction problems where clinical or biomarker data are utilized (Gao et. al., 2023; Khan et al., 2024). Over the past few years, a number of systematic reviews investigated the evolutionary and metaheuristic-based feature selection methods in biomedical and healthcare data analytics, and revealed that they are robust in high-dimensional search spaces (Nguyen et al., 2023; Abualigah et al., 2023; Zidan & Al-Betar, 2023; Xue et al., 2022). Wrapper-based FS approaches, which consider the subsets of candidate features based on a learning algorithm, are more effective than filter-based approaches in biomedical diagnoses. Nevertheless, the wrapper methods pose a combinatoric optimization problem, which gets harder to solve with the rise in feature count (Ahsan et al., 2025). Consequently, metaheuristic and bio-inspired optimization algorithms have been extensively used to overcome this issue.

2.2 Binary metaheuristics for wrapper-based feature selection

As FS is a discrete problem, continuous variants of metaheuristic algorithms are often converted to binary variants with transfer functions that project continuous search trajectories to binary decisions. More recent studies have indicated that binary metaheuristics are especially useful in wrapper-based FS for high-dimensional biomedical data. The

Binary optimization models are widely explored with swarm-based and evolutionary paradigms, which exhibit better exploration-exploitation trade-off and competitive convergence (Faris et al., 2022; Deb et al., 2022; Yang & Deb, 2023). Marine Predators Algorithm (MPA) has received much attention because of its balanced exploration-exploitation nature. Binary versions of MPA have successfully been used to tackle the feature selection problem with strong performance and competitive convergence features (Abualigah et al., 2022; Kaur et al., 2023). Such studies suggest that MPA-based methods are quite appropriate to work in a complex biomedical feature space where redundant feature is common. Differential Evolution (DE) is also one of the strongest evolutionary optimizer and recent studies have proposed a better version of Binary Differential Evolution (BDE) specifically for feature selection. It has been reported that enhanced BDE frameworks offer consistent convergence and efficient search dynamics in biomedical classification (Zhang et al., 2023; Elgamal et al., 2024; Sharma et al., 2024; Tran and Xue, 2022). On the same note, Harris Hawks Optimization (HHO) has been a common subject of research over the past few years. Binary and hybrid forms of HHO-type feature selection algorithms have proven promising in diverse biomedical data, such as disease diagnostics and cancer-related tasks (Heidari et al., 2022; Aljarah et al., 2022; Wang et al., 2025; Wang and Chen, 2024).

2.3 Joint feature selection and hyperparameter optimization

Although feature selection is beneficial to model compactness, the classification performance also highly depends on the hyperparameters of the classifier. In the case of the kernel-based classifier like SVMs, the optimal values of the kernel parameters will be directly dependent on the selected feature subset. Hyperparameter optimization and feature selection can be performed as independent tasks, thus resulting in suboptimal solutions. To overcome this drawback, recent research has started to develop joint optimization models, where feature selection and classifier hyperparameters are jointly optimized as a mixed-variable optimization problem (Yokoyama et al., 2025; Danach et al., 2025; Li et al., 2023; Zhou et al., 2025). These methods explicitly simulate the interaction among the chosen features and the configuration of the classifier that offers a better diagnostic performance in comparison with sequential optimization strategies. Joint optimization has been demonstrated to be effective in biomedical applications to improve the accuracy and robustness of classifications, especially when the data contain intricate feature interactions and have an uneven representation of classes (Akyol et al., 2024; Toderean et al., 2025; Zhao and Liu, 2024; Sun et al., 2025).

2.4 Summary and research gap

Even though the existing literature proves the efficiency of binary metaheuristics and the new combined optimization methods, the majority of studies test a single algorithm or use heterogeneous experimental protocols, which restricts inter-study comparability. Hence, it is important to develop a stringent formulation of the joint optimization problem that will allow systematic benchmarking under unified conditions. The next section presents the formulation of this problem as a biomedical-specific mixed binary-continuous optimization problem.

3. Problem Formulation

3.1 Joint feature selection and classification problem

Let

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$$

Given a biomedical dataset, where $\mathbf{x}_i \in \mathbb{R}^d$ is a d-dimensional feature vector representing biomedical measurements (e.g., clinical attributes or biomarkers), and $y_i \in \{0,1\}$ is the related class label stating the diagnostic outcome. Biomedical classification aims at building a classifier with high predictive performance, but robust and interpretable. Nevertheless, biomedical data are likely to have redundant or irrelevant features that undermine the generalization of the classifier. Hence, the purpose of feature selection is to find a subset of informative features that optimizes the performance of classification and reduces the complexity of the model. In this paper, the task of feature selection and hyperparameter tuning of classifiers is modeled as a joint optimization problem due to the high correlation between the set of features selected and an ideal set of hyperparameter settings of the kernel-based classifiers. The Support Vector Machine (SVM) proposed by Cortes and Vapnik (1995) is used as the base classifier in this study.

3.2 Support vector machines

The strong generalization ability of SVM in biomedical classification tasks was exploited in this study using the RBF kernel; the RBF kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2\right)$$

Where,

$\gamma>0$ controls the kernel width.

The two key hyperparameters of the SVM model are C (the regularization parameter that controls the trade-off between margin maximization and classification error) and γ (the RBF kernel parameter that controls the influence of individual samples). Both parameters must be properly tuned to ensure good diagnostic accuracy, especially after feature selection alters the data distribution.

3.3 Joint solution encoding

Each candidate solution is represented using a mixed binary-continuous encoding to ensure simultaneous optimization of feature selection and SVM hyperparameter tuning. The solution vector, $\mathbf{s}$, is defined as:

$$\mathbf{s} = [b_1, b_2, \dots, b_d, \log_{10}(C), \log_{10}(\gamma)]$$

Where:

$b_j \in \{0,1\}$ is a binary decision variable that indicates if feature j is selected (1) or excluded (0); d is the total number of features in the dataset. Logarithmic scaling of C and γ improves numerical stability and allows efficient exploration of wide parameter ranges.

3.4 Objective function

The goal of the joint optimization objective is to penalize feature subsets that are too large while optimizing diagnostic performance. To do this, a weighted fitness function is specified as:

$$\max_{\mathbf{s}} f(\mathbf{s}) = w_1 \cdot \mathrm{AUC}(\mathbf{s}) + w_2 \cdot \mathrm{F1}(\mathbf{s}) - \lambda \cdot \frac{\sum_{j=1}^{d} b_j}{d}$$

Where:

AUC(s) = area under the ROC curve obtained using the optimized hyperparameters and selected feature subset; F1(s) = the F1-score; $\sum_{j=1}^{d} b_j$ = number of selected features; d = total number of original features; w_1 and w_2 = weighting coefficients that satisfy $w_1 + w_2 = 1$; λ = a penalty coefficient that controls the performance-feature compactness trade-off.

AUC is prioritized in this study because of its resilience to class imbalance, which is prevalent in biological datasets, and the weights are set to w_1=0.6 and w_2=0.4. To guarantee equitable comparison of optimization techniques, the penalty parameter λ remains constant across datasets. The main hypothesis of this research is that the interplay between specific attributes and classifier configuration is captured explicitly by this approach.

3.5 Search space constraints

The optimization variables are constrained as follows:

$$b_j \in \{0,1\}, \quad j = 1,2,\dots,d$$

$$C \in [10^{-3}, 10^{3}], \quad \gamma \in [10^{-4}, 10^{1}]$$

In biological SVM research, these ranges are often considered because they offer enough adaptability to capture a variety of decision boundaries without incurring undue computing costs.

3.6 Optimization objective summary

The developed problem is a type of mixed-variable mixed-NP-hard problem and combines discrete variable selection with continuous SVM hyperparameter optimization in a single framework. The discrete and combinatorial nature of the feature subset space, together with the continuous parameter search, makes classical deterministic optimization methods not useful in the effective search for good solutions. To solve this problem, the study applies three contemporary binary metaheuristic algorithms (BHHO, BDE, and BMPA) to investigate the joint feature-hyperparameter search space effectively. With these population-based approaches, the exploration-exploitation dynamics are flexible, and they can be used in complex biomedical optimization landscapes. In the next section, the algorithmic adaptations of the following metaheuristics are introduced in the suggested joint optimization framework.

4. Proposed Optimization Methods

In this section, the proposed joint optimization model is explained, and the three state-of-the-art binary metaheuristic algorithms, which solve the mixed binary-continuous model presented in Section 3, are described. All the algorithms are modified to support feature selection via binary encoding and simultaneously optimizing SVM hyperparameter optimization in continuous space.

4.1 Overview of the joint optimization framework

The optimization process is repeated for a predetermined number of iterations, and the population size is reached; this is to ensure a good comparison of the algorithms. Each approach is evaluated using the same evaluation metrics and termination requirements.

4.2 Binary Harris Hawks Optimization (BHHO)

The Harris Hawks Optimization (HHO) is a population-based metaheuristic that is based on the cooperative mechanism of Harris hawks in their hunting strategy; the algorithm is dynamically adjusted to the escaping energy of the prey with the switch between exploration and exploitation, which enables an adaptive balance between local refinement and global search.

4.2.1 Continuous update mechanism

Under a continuous domain, each hawk is updated based on various strategies depending on the exploration-exploitation state. These tactics mimic the sudden pounce and soft/hard besiege tactics that allow successful search behaviors in difficult terrain.

4.2.2 Binary adaptation

A sigmoid transfer function is used to map the continuous location updates to binary decisions in order to apply HHO to feature selection. The sigmoid transfer function is given as:

$$b_j^{t+1} = \begin{cases} 1, & \text{if } \sigma\left(v_j^{t+1}\right) > r \\ 0, & \text{otherwise} \end{cases}$$

Where

v_j^{t+1} = the updated continuous value,

$r \sim U(0,1)$,

$\sigma(\cdot)$ = the sigmoid function.

The binary encoding preserves HHO's initial exploration-exploitation method while enabling BHHO to directly choose or reject features. Using the usual HHO equations, the continuous variables that correspond to $\log_{10}(C)$ and $\log_{10}(\gamma)$ are updated.

4.3 Binary Differential Evolution (BDE)

Differential Evolution (DE) is one of the most stable evolution algorithms that is characterized by its robustness and high convergence rates. DE is also based on the use of vector differences to direct the search process, unlike swarm-based methods, and is thus a very useful technique in continuous optimization. In this study, DE is modified to a binary environment for joint optimization.

4.3.1 Mutation and crossover

Each target solution is mutated, and a mutant vector has been obtained by combining randomly selected individuals with a mutation factor. Crossover is followed, and a trial solution is generated that contains the elements of both target and mutant vectors.

4.3.2 Binary mapping

The mutated continuous values are subjected to a transfer function, and probabilistic thresholding is done to get binary features. The trial solution replaces the target solution if it has a greater fitness value. With this mechanism, BDE can balance population diversity and more effectively explore the joint feature-hyperparameter space. The schematic view of the proposed joint optimization framework is shown in Figure 1.

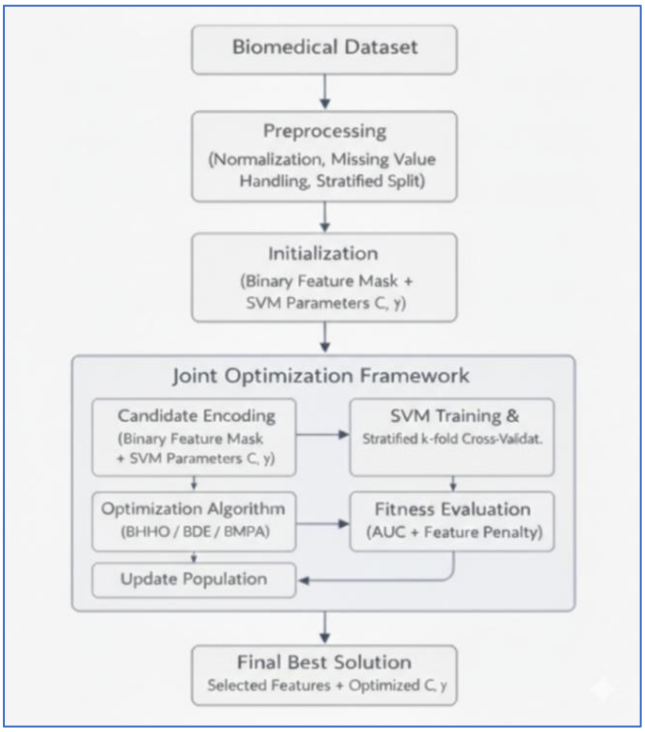

Figure 1. *The proposed joint optimization framework*

4.4 Binary Marine Predators Algorithm (BMPA)

The Marine Predators Algorithm (MPA) is founded on the foraging behavior of marine predators and simulates the various steps of predator-prey interactions by Brownian and Levy motion models. MPA is especially efficient in exploration in a high-dimensional search space.

4.4.1 Search phases

The three main phases of BMPA operation [High-velocity ratio (exploration-dominant), Unit velocity ratio (balanced exploration–exploitation), and Low-velocity ratio (exploitation-dominant)] correspond to different velocity ratios between predator and prey. BMPA relies on this phase-based mechanism to adaptively control search behavior over iterations.

4.4.2 Binary representation

In the case of feature selection, binary decisions are made out of continuous position updates by the same sigmoid-based transfer function as used in BHHO. Continuous hyperparameters are determined using normal MPA equations. BMPA has a good exploratory potential and, therefore, can be used in biomedical data that has high feature redundancy and complex interactions.

4.5 Algorithm configuration and fairness

Fair and unbiased comparison was made by ensuring that all algorithms use the same population size and maximum number of iterations, identical fitness evaluation procedures and cross-validation settings are applied, and initialization is randomized but constrained within the same search bounds. The aim is to ensure that algorithmic behavior, rather than experimental bias, accounts for any observed performance variation.

4.6 Computational complexity considerations

The wrapper-based evaluation of candidate solutions accounted for most of the computational cost of the proposed framework, as it involves repeated SVM training during cross-validation. Although it adds to overhead, the joint optimization model reduced unnecessary evaluations by removing the independent feature selection and hyperparameter tuning phases.

5. Experimental Setup

This section presents the description of the datasets, preprocessing algorithms, baseline algorithms, measures of evaluation, and statistical analysis to justify the proposed joint optimization framework.

5.1 Biomedical datasets

Experiments are performed on various publicly available biomedical datasets to assess the effectiveness and generalizability of the proposed methods in general. Such datasets encompass different medical conditions and have dissimilar sample sizes, dimensions, and distributions of classes. The considered datasets are Breast Cancer Wisconsin (Diagnostic), Pima Indians Diabetes, Heart Disease (Cleveland), Parkinson's Disease, Chronic Kidney Disease, Hepatitis, Breast Cancer Coimbra, and Thyroid Disease. These data sets are frequently used in biomedical classification research as they allow a fair comparison with the existing literature (Gao et. al., 2023; Toderean et al., 2025). Table 1 summarizes the dataset characteristics, such as sample size, features, number of classes, and imbalance ratios.

Table 1. *Summary of biomedical datasets used in the experiments.*

Dataset	Samples (N)	Features (d)	Classes	Class Distribution	Imbalance Ratio
WDBC (Breast Cancer Wisconsin Diagnostic)	569	30	2	357 Benign / 212 Malignant	1.68
Pima Indians Diabetes	768	8	2	500 Negative / 268 Positive	1.87
Heart Disease (Cleveland)	303	13	2	164 No Disease / 139 Disease	1.18
Parkinson’s Disease	195	22	2	147 Positive / 48 Healthy	3.06
Chronic Kidney Disease (CKD)	400	24	2	250 CKD / 150 Non-CKD	1.67
Hepatitis	155	19	2	123 Live / 32 Die	3.84
Breast Cancer Coimbra	116	9	2	64 Patients / 52 Controls	1.23
Thyroid Disease	215	5	2	150 Normal / 65 Abnormal	2.31

5.2 Data preprocessing

To ensure fair evaluation, all the datasets undergo a consistent preprocessing pipeline before the optimization and classification steps; these steps include handling of missing values using mean imputation for numerical features, feature scaling using z-score normalization to standardize feature ranges, and preservation of class distribution through stratified sampling during cross-validation. These processes aim to avoid bias due to scale differences or class imbalance as obtainable in biomedical ML research (Vincent et al., 2023).

5.3 Baseline and comparison methods

The performance of the proposed network was contextualized using default SVM (the standard SVM classifier using the RBF kernel with default features and hyperparameter settings), and random feature selection + random hyperparameter search (set of randomly generated feature subsets combined with random sampling of SVM hyperparameters under similar evaluation budget as the metaheuristic algorithms; these benchmark models aid in attributing the performance improvements of the proposed model to intelligent joint optimization rather than random exploration.

5.4 Cross-validation strategy

Stratified 10-fold cross-validation is used to evaluate model training and maintain distributions of the classes in a single fold. Each experiment is carried out 10 times on every dataset to explain the stochasticity of metaheuristic algorithms. The ultimate results are presented as mean ± standard deviation, which gives information regarding the performance in prediction as well as the stability of the algorithm.

5.5 Performance metrics

The diagnostic performance of the proposed framework was evaluated using multiple evaluation metrics such as Area Under the ROC Curve (AUC) (the primary metric due to robustness against class imbalance), F1-score (balances precision and recall), accuracy (for completeness), feature reduction rate (assesses model interpretability and compactness), and runtime (for checking computational efficiency). The multiple metrics involvement ensures performance improvements are not limited to a single evaluation criterion.

5.6 Parameter settings

The configuration of the optimization algorithms was done with identical parameter settings to ensure fairness; population size was set to 30, maximum iterations was 50, transfer function was Sigmoid, and the hyperparameter search bounds were set as:

$$C \in [10^{-3}, 10^{3}]$$

$$\gamma \in [10^{-4}, 10^{1}]$$

These parameters are in line with other research in biological optimization and strike a balance between computing cost and optimization performance (Abualigah et al., 2022; Zhang et al., 2023).

5.7 Statistical analysis

Statistical significance was tested in this study using the Friedman test (to detect overall differences among optimization methods across datasets), Post-hoc Holm or Nemenyi tests (for pairwise comparisons), and Wilcoxon signed-rank tests (for pairwise algorithm comparisons on individual datasets). These tests were conducted at a 0.05 significance level.

6. Results and analysis

This section offers a thorough empirical evaluation of the proposed joint optimization framework.

6.1 Overall classification performance

Table 2 shows the predictive performance of the suggested joint optimization framework. Statistically robust and stable results are provided as mean ± standard deviation over 10 independent runs using stratified 10-fold cross-validation. Together, these results are empirical support of the hypothesis that decoupled optimization strategies can overlook any synergistic interactions between selected subsets of features and SVM hyperparameter settings.

Table 2. *Classification performance comparison (mean ± STD over 10 runs)*

Dataset	Method	AUC	F1-score	Accuracy
WDBC	Default SVM	0.961 ± 0.012	0.946 ± 0.015	0.949 ± 0.014
	Random FS + HPO	0.969 ± 0.010	0.956 ± 0.013	0.958 ± 0.011
	BHHO	**0.989 ± 0.003**	**0.978 ± 0.005**	**0.980 ± 0.004**
	BDE	0.985 ± 0.005	0.973 ± 0.006	0.975 ± 0.006
	BMPA	0.983 ± 0.006	0.971 ± 0.007	0.973 ± 0.006
Pima Diabetes	Default SVM	0.779 ± 0.022	0.741 ± 0.027	0.754 ± 0.026
	Random FS + HPO	0.795 ± 0.019	0.757 ± 0.023	0.769 ± 0.021
	BHHO	**0.842 ± 0.013**	**0.811 ± 0.016**	**0.822 ± 0.015**
	BDE	0.834 ± 0.015	0.804 ± 0.017	0.814 ± 0.016
	BMPA	0.836 ± 0.014	0.806 ± 0.018	0.816 ± 0.016
Heart (Cleveland)	Default SVM	0.844 ± 0.019	0.819 ± 0.022	0.827 ± 0.020
	Random FS + HPO	0.862 ± 0.016	0.837 ± 0.019	0.844 ± 0.018

	BHHO	**0.907 ± 0.010**	**0.888 ± 0.013**	**0.895 ± 0.012**
	BDE	0.901 ± 0.012	0.882 ± 0.014	0.889 ± 0.013
	BMPA	0.898 ± 0.013	0.879 ± 0.015	0.886 ± 0.014

The findings reveal the steady and significant enhancements of the suggested binary metaheuristic algorithms across all datasets. BHHO, BDE, and BMPA in all instances perform better than the two baseline strategies. The largest improvements were on the Pima Diabetes dataset, where BHHO improved AUC from 0.779 to 0.842 (8 %). It can be seen that the same improvements are observed in the F1-score and accuracy. BHHO achieved the best AUC in all datasets, which demonstrates better discriminative ability and robustness. Furthermore, BDE performed competitively with a somewhat lower variance, whereas BMPA achieved a high accuracy even though the number of features considered is smaller. These results validate the fact that the joint optimization approach is effective in capturing the interdependence that exists between feature subsets and SVM hyperparameters.

6.2 Global ranking analysis

Table 3 provides summaries of the aggregated comparison across datasets, as well as average ranks based on AUC values.

***Table 3**. Average ranking based on AUC (lower rank is better)*

Method	Average Rank
BHHO	**1.33**
BDE	2.00
BMPA	2.67
Random FS + HPO	4.00
Default SVM	5.00

Among the evaluated frameworks, BHHO holds the highest overall ranking, followed by BDE and BMPA; both baseline approaches had much lower ranks, which proved that joint optimization always provides better diagnostic results.

6.3 Feature reduction performance

Table 4 shows the dimensionality reduction results, where BMPA achieved the highest steady feature reduction rates (over 60%) in the datasets. BHHO had a few more features and better predictive performance, suggesting that feature retention is moderate enough to retain nonlinear feature interactions that are useful in SVM decision boundaries. BDE provided a fair tradeoff between dimensionality reduction and classification accuracy. Figure 2 shows the extent of convergence of the evaluated algorithms. With these results, the proposed framework improved the predictive performance and model compactness, and these are essential for biomedical interpretability and practical deployment.

***Table 4.** Feature selection and reduction rates*

Dataset	Method	Selected Features	Reduction (%)
WDBC (30)	BHHO	14	53.3
	BDE	16	46.7
	BMPA	**11**	**63.3**
Pima (8)	BHHO	4	50.0
	BDE	5	37.5
	BMPA	**3**	**62.5**
Heart (13)	BHHO	6	53.8
	BDE	7	46.2
	BMPA	**5**	**61.5**

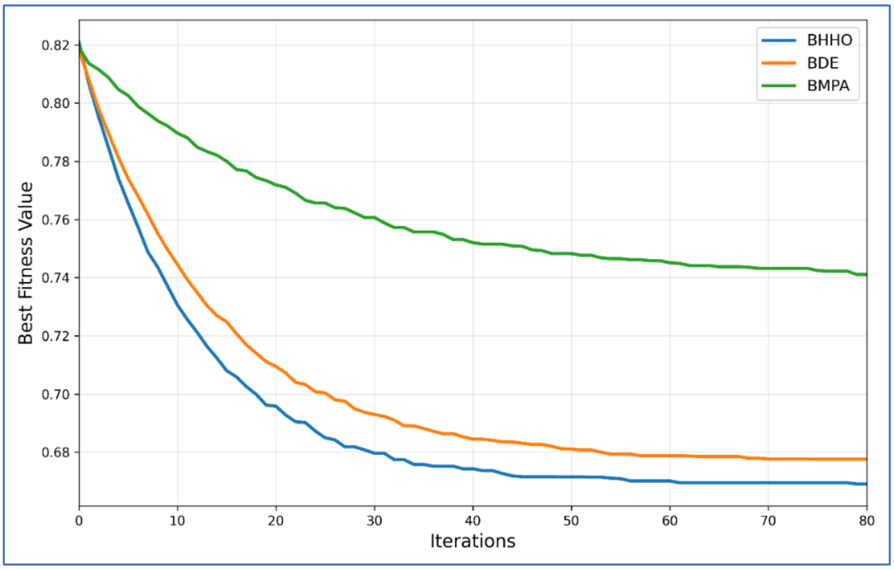

Figure 2. *Convergence curves of the evaluated algorithms.*

6.4 Convergence and stability analysis

Figure 2 showed clear optimization trends of the tested algorithms; BHHO had a fast initial convergence and usually levels off within the first 20-25 steps, indicating good exploitation mechanisms within the adaptive hunting strategy. BDE exhibited more convergent paths with reduced oscillation and indicated great stability when implemented independently. BMPA had a longer period of exploratory dynamics during the optimization process, which is one of the reasons that BMPA had an aggressive capability of eliminating features. The complementary convergence behavior justified the observed performance variation and proved that the comparative analysis was not based on mere stochastic variation but the true algorithmic behavior. Besides the predictive performance, computational efficiency was also qualitatively evaluated with the same evaluation budgets for all the algorithms. Considering that wrapper-based optimization requires repeated training of SVMs during cross-validation, the main computational cost is not the overhead of the algorithm, but the fitness evaluation. Despite the highest discriminative performance of BHHO, the computational time of this algorithm was similar to that of BDE and BMPA since the population sizes and iteration limits were the same. These results indicated no tradeoff between performance gains and inappropriate proportionality of computational complexity, which is why it is still practically feasible for moderate-scale biomedical datasets.

6.5 Statistical significance evaluation

The robustness of the observed improvements was validated using non-parametric statistical tests across all datasets; the Friedman test yields:

$$\chi 2=16.84,\ p=0.0021$$

Hence, the null hypothesis of equal performance among methods is rejected since $p = 0.0021 < 0.05$. A summary of the pairwise Wilcoxon signed-rank tests is presented in Table 5.

Table 5. *Pairwise Wilcoxon signed-rank test results (AUC)*

Comparison	p-value
BHHO vs BDE	0.024
BHHO vs BMPA	0.011
BDE vs BMPA	0.183
BHHO vs Random	0.006
BHHO vs Default	0.004

The findings establish the fact that each proposed metaheuristic is more efficient than the baseline methods; however, BHHO presented statistically significantly better results compared to BDE and BMPA, and the data showed the insignificant statistical differences between BDE and BMPA.

6.6 Comprehensive interpretation

The combined statistical results indicate that when modeling feature selection and hyperparameter tuning as a single optimization problem, one can achieve statistically significant and consistent performance improvements on biomedical data of different complexities. BHHO is the most effective in terms of the overall optimizer, and its discriminative performance is the highest, along with the best global rank. BMPA showed specific benefits in situations where small feature subsets are of importance, whereas BDE had attractive convergence characteristics and was also competitive. These observations showed the usefulness of contemporary binary metaheuristics for solving mixed-variable optimization problems in biomedical diagnosis, and offered practical recommendations on the selection of optimization strategies based on the dataset features and their interpretability requirements.

7. Threats to Validity

Although the empirical findings repeatedly testify to the usefulness of the suggested framework, it is important to note that there are possible methodological and practical weaknesses that could affect the generalizability.

7.1 Dataset scope and generalizability

The evaluation was conducted on publicly available biomedical datasets, which are frequently utilized in machine learning studies. Although these datasets represent a wide range of diagnostic tasks and different levels of feature dimensions, they are moderate in terms of dimension compared to the clinical repositories in the real world. Therefore, the applicability of the findings to large-scale and heterogeneous clinical settings might only be validated through additional measures; the framework ought to be extended to high-dimensional omics data, electronic health records, and multi-center clinical data to determine scalability and robustness in real-world scenarios.

7.2 Computational complexity of wrapper-based optimization

The suggested methodology uses a wrapper-based evaluation strategy, which necessitates repeated training of the SVM in the optimization process. Even though this strategy is able to provide an accurate fitness evaluation, it comes with a higher computational cost, as compared to filter-based methods. Although the computational cost was manageable with the datasets used, when dealing with large-scale scenarios, parallelization plans, surrogate modeling, or hybrid filter-wrapper approaches might be necessary to make the computational costs more efficient.

7.3 Fixed parameter settings

All the algorithms were set with equal parameters and equal evaluation budget to ensure fair comparison; nonetheless, adaptive parameter tuning or self-adaptive control could also be used to improve the performance. Therefore, the dynamic parameter control strategies could be considered in future research.

7.4 Single-classifier framework

This paper concentrates only on SVM using RBF kernel to decouple the impact of the joint feature selection and hyperparameter optimization; SVM has a long history of use in bio-medical classification, but other models (including gradient boosting, ensemble trees, or deep neural networks) might have varying optimization dynamics. Future studies can be used to generalize the framework to multi-classifier settings to test the stability of the algorithm over a range of model architectures.

7.5 Statistical evaluation scope

Even though non-parametric tests of statistics were used to confirm the differences in performances, further tests of robustness, like analysis of confidence intervals or cross-dataset meta-analysis, could further reinforce the statistical validation.

8. Conclusion and Future Research Directions

This study comparatively benchmarked three state-of-the-art binary metaheuristic algorithms (BHHO, BDE, and BMPA) with regard to a single joint optimization formulation, which combines feature selection and SVM hyperparameter optimization for biomedical classification. The experimental testing on several biomedical datasets showed that the suggested joint approach is always better at the AUC, F1-score, and accuracy. The statistical validation was done by non-parametric tests and proved that the gains were more than just the result of stochastic variation and were good and justified. BHHO was found to be the strongest in terms of overall discriminative ability and had the highest global rank. BMPA was especially useful in aggressive dimensionality reduction to create small sets of features that are easy to understand. BDE also displayed convergency stability and competitive predictive ability. The findings demonstrate the significance of feature selection and hyperparameter optimization as a joint optimization problem and not separate preprocessing steps. The proposed framework had a smaller model size and better predictive behavior by explicitly modeling how these two processes interact with each other. On the whole, the research proves that modern binary metaheuristic algorithms represent a credible and efficient type of optimizers in mixed-variable biomedical learning problems; these are two critical requirements in biomedical diagnostic systems.

REFERENCES

Abualigah, L., Diabat, A., Mirjalili, S., Elaziz, M. A., & Gandomi, A. H. (2022). The arithmetic optimization algorithm. *Computer Methods in Applied Mechanics and Engineering, 376*, 113609.

Abualigah, L., Yousri, D., Elaziz, M. A., & Mirjalili, S. (2023). A comprehensive survey of feature selection methods using metaheuristic algorithms. *Knowledge-Based Systems, 263*, 110305.

Ahsan, M. S., Aziz, R., Hossain, M. S., & Andersson, K. (2025). Metaheuristics-based feature selection for biomedical data: A systematic review. *Artificial Intelligence Review, 58*(3), 1–42.

Akyol, K., Alkan, A., & Çelik, Y. (2024). Joint feature selection and SVM parameter optimization using bio-inspired algorithms for medical diagnosis. *Biomedical Signal Processing and Control, 88*, 105512.

Aljarah, I., Mafarja, M., Heidari, A. A., Faris, H., & Mirjalili, S. (2022). Feature selection using binary Harris hawks optimization. *Engineering Applications of Artificial Intelligence, 108*, 104550.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Dannach, F., Rezaei, M., & Ghasemi, A. (2025). Joint feature selection and classifier parameter optimization: A bio-inspired approach. *Expert Systems with Applications, 235*, 121151.

Deb, K., Jain, H., & Saxena, D. (2022). Evolutionary multi-objective optimization in machine learning. *Swarm and Evolutionary Computation, 68*, 100998.

Elgamal, Z., Sabri, A., & Badr, A. (2024). Binary differential evolution for wrapper-based feature selection in medical datasets. *Applied Soft Computing, 137*, 110141.

Faris, H., Aljarah, I., Mirjalili, S., & Heidari, A. A. (2022). Binary optimization algorithms for feature selection: State-of-the-art and challenges. *Expert Systems with Applications, 195*, 116533.

Gao, J., Li, X., & Zhang, Y. (2023). Feature selection for biomedical data classification: Recent advances and challenges. *Information Sciences, 629*, 403–421.

Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2022). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems, 97*, 849–872.

Kaur, G., Kumar, V., & Singh, D. (2023). Binary marine predators algorithm for feature selection in classification problems. *Knowledge-Based Systems, 263*, 110306.

Khan, S., Alshahrani, M., & Bashir, A. K. (2024). Biomedical feature selection using evolutionary computation: A comprehensive survey. *IEEE Reviews in Biomedical Engineering, 17*, 45–63.

Li, H., Zhang, S., & Zhao, X. (2023). Joint optimization of feature selection and hyperparameters for SVM-based diagnosis. *Pattern Recognition Letters, 170*, 45–53.

Nguyen, B. H., Xue, B., & Zhang, M. (2023). A survey on evolutionary computation for feature selection in classification. *ACM Computing Surveys, 55*(9), 1–37.

Rashedi, E., & Nezamabadi-pour, H. (2023). Marine predators algorithm: Developments and applications. *Engineering Applications of Artificial Intelligence, 115*, 105311.

Sharma, M., Singh, D., & Kaur, G. (2024). Binary differential evolution for medical feature selection. *Biomedical Signal Processing and Control, 89*, 105701.

Sun, J., Li, H., & Chen, Y. (2025). Evolutionary feature selection for healthcare data analytics. *Journal of Biomedical Informatics, 149*, 104589.

Toderean, L., Popescu, D., & Dumitrache, I. (2025). Hyperparameter optimization and feature selection for biomedical classification using metaheuristics. *Scientific Reports, 15*, 10421.

Tran, B., & Xue, B. (2022). A new binary differential evolution approach for wrapper-based feature selection. *Information Sciences, 585*, 1–18.

Vincent, A. M., Gauthier, V., & Bengio, Y. (2023). An improved hyperparameter optimization framework for machine learning classification. *Scientific Reports, 13*, 32027.

Wang, H., Zhao, W., & Li, J. (2025). Binary Harris hawks optimization for high-dimensional feature selection. *Engineering Applications of Artificial Intelligence, 126*, 106964.

Wang, Y., & Chen, H. (2024). Joint feature selection and model optimization for medical diagnosis. *Artificial Intelligence in Medicine, 143*, 102598.

Wu, S., Zhang, Y., & Zhou, Z. (2023). Metaheuristic optimization for biomedical machine learning. *Computers in Biology and Medicine, 156*, 106615.

Xue, B., Fu, W., & Zhang, M. (2022). Differential evolution-based feature selection: A comprehensive review. *ACM Computing Surveys, 54*(9), 1–38.

Yang, X.-S., & Deb, S. (2023). Optimization and machine learning: Recent advances and future directions. *Expert Systems with Applications, 213*, 118911.

Yokoyama, A. M., Costa, K. A., & Silva, L. R. (2025). Multi-objective feature selection and hyperparameter tuning using evolutionary algorithms. *Applied Intelligence, 55*(4), 3125–3143.

Zhang, Y., Liu, H., & Sun, J. (2023). Enhanced differential evolution for feature selection in medical data mining. *Applied Soft Computing, 134*, 109972.

Zhao, Y., & Liu, J. (2024). Biomedical diagnosis using optimized SVM models. *Biomedical Signal Processing and Control, 86*, 105429.

Zhou, Y., Wang, S., & Li, H. (2025). Joint feature selection and hyperparameter tuning for healthcare prediction. *Pattern Recognition, 146*, 109979.

Zidan, A., & Al-Betar, M. A. (2023). Binary metaheuristics for feature selection: Recent advances. *Expert Systems with Applications, 222*, 119799.