



Efficient hybrid optimization based feature selection and classification on high dimensional dataset

A. Ameer Rashed Khan¹ · S. Shajun Nisha²

Received: 19 September 2022 / Revised: 12 October 2023 / Accepted: 21 November 2023 /
Published online: 23 December 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

With the vast usage of intelligent information systems (ISs), the tremendous increase in data volume creates numerous problems and challenges, such as high dimensionality, noisy and irrelevant data. These issues lead to high computational costs and greatly affect the accuracy and efficiency of machine learning (ML) algorithms. Feature selection (FS) is one of the most important concepts used effectively to boost the classification's accuracy and minimize computational costs. However, finding an effective FS approach is challenging, and numerous swarm-based algorithms inspired by biological systems have been developed. Feature selection aims to determine the best subset of features for categorizing the class labels by eliminating irrelevant data. This paper introduces the hybrid optimization approach to solve the problems in the feature selection process. The input data is obtained from several datasets, and the data cleaning is performed in the pre-processing stage. Initially, eight different optimization techniques are executed and depending on the results attained from performance metrics, the best two algorithms are selected. The selected best algorithms are combined together to generate a hybrid process. The proposed work hybridizes a Slime Mould Algorithm (SMA) with Binary Grey Wolf Optimization (BGWO) for feature selection. The selected features from the hybrid algorithms are fed to the K-nearest neighbor (KNN) classifier, which is analyzed to be effective compared to the other classifier. Finally, the hybrid SMA + BGWO based feature selection with the KNN classifier effectively solves the FS problems on high dimensional data with remarkable accuracy and convergence speed. Performance metrics like accuracy, precision, F-measure, computational time, recall, RMSE and MAE are utilized to evaluate the efficacy of classifiers. The proposed SMA + BGWO approach with KNN classification in the CICDDoS2019 dataset attains an accuracy of 99.83%, and CICMalDroid2020 attains an accuracy of 99.30%. The experimental analysis proves that the proposed hybrid technique is better than the existing techniques.

Keywords Optimal feature selection · Swarm models · High dimensionality data · Slime mould algorithm · Binary grey wolf optimization · K-nearest neighbor

1 Introduction

Feature selection is an essential task that can disrupt the efficacy of image recognition and classification [1]. The major intention of the feature selection process is to minimize the large dimensionality of data that affects classification performance. Minimizing the dimensionality of patterns through the feature selection process is the most challenging task for image classification [2]. The feature selection process applies to several fields, such as signal processing, data mining, bioinformatics, text classification, pattern recognition, image processing, etc. This feature selection stage selects a subset of the presented features by ignoring unwanted features [3]. Feature dimensionality reduction is necessary to attain better information in a given image set. This process helps to diminish the redundant features, which are highly correlated, also reducing the overall computational time [4].

Recently, various approaches have been developed to address the issues of minimizing irrelevant features in image processing fields [5]. An effective feature selection technique assists in learning data, diminishing the demand for computation, minimizing the dimensionality curse effect, and enhancing classification performance. From the point of machine learning techniques, if a system utilizes unwanted features, it will utilize such information for advanced data, leading to the worst generalization. Eliminating unwanted features must not be examined with other size reduction approaches like Principal Component Analysis (PCA) [6] since optimal features can be unique to the remaining dataset. The feature reduction process cannot generate additional features since it utilizes the input features to diminish the number of features. When the features are chosen, a concept must be established to identify the subset of optimal features. If the feature subsets are measured directly for provided data, an NP-hard issue arises as the amount of features increases. Therefore, a suboptimal concept must be utilized to ignore high data dimensionality with complex computations [7].

In general, the dataset involves large amounts of information with unwanted or irrelevant data, which can reduce the machine learning model's performance. This problem can be overcome by selecting the optimal set of highly suitable features for classification [8]. The applications of feature selection are carried out in varied areas like face recognition, classification of genes, plant disease classification, classification of tumor, etc. The feature selection is assumed to be an NP-complete combinatorial optimization issue, which intends to choose n optimal amount of features from the overall amount of features m without ignoring the information. Thus, various optimization algorithms are introduced to resolve the feature selection issues, such as genetic algorithm (GA) [9], Particle swarm optimization (PSO) [10–12], binary dragonfly algorithm (BDA) [13–15], whale optimization algorithm (WOA) [16, 17], etc. This problem can be solved using filtering, hybridized, wrapper, and embedded methods. The filtering based approaches cannot require learning models, and these approaches utilize statistical models to compute a subset of features or find the feature correlations [18]. The wrapper approaches utilize learning models, including different classifiers, to compute a subset of features. Compared with wrapper approaches, the filter methods are increased in speed. However, the wrapper based techniques attain improved classification accuracy than filter approaches [19, 20]. Some of the wrapper techniques employed for feature selection are Naïve Bayes (NB), Decision tree (DT), KNN, linear discriminant analysis (LDA) and so on.

1.1 Motivation

The feature selection effectively finds the most important features from the dataset and removes redundant (irrelevant) data. The main goals of feature selection include a better understanding of the data for different ML applications, maximizing prediction performance, and reducing data dimensionality. The data set, which consists of redundant and noisy data features, can significantly slow down the processing speed of the learning algorithm and affect its accuracy. High-dimensional data affects computation time, learning algorithm, model accuracy, and computational resources (memory). Until now, the classical optimization techniques for solving feature selection problems have been impractical. Therefore, evolutionary algorithms (EAs) are developed as an alternative to finding the optimal solution. In a group, EAs are generally inspired by nature, the biological and social behavior of birds, animals, fish, wolves, bats, fireflies, etc. Many researchers have offered various calculation methods to mimic the behavior of these species and find an optimal solution. Therefore, reducing the dimensionality of data and solving its problem is crucial to providing the most flexible, reliable, and highly accurate computational applications [21]. This motivated the authors to develop a hybrid optimization algorithm for feature selection that can shorten the learning time, reduce the size of features, and/or improve the performance of the classification algorithm.

1.2 Contribution

The major contributions of this work are as follows,

- To solve the problems of high-dimensional data sets, a new hybrid model for optimal feature selection is developed, and the algorithm performance is improved.
- To analyze multiple optimization algorithms with the high-dimensional dataset in order to choose the first and second best algorithms to form the finest hybrid combination.
- The proposed work develops an optimized hybrid technique using SMA and BGWO algorithms to make the feature selection process more effective.
- To efficiently remove the redundant or irrelevant data features, allow accurate search analysis, and speed up the algorithm with increased computational efficacy and high accuracy.

The rest of this paper is structured as follows: Section 2 describes the literature survey related to the proposed work, Section 3 deals with the proposed methodology, Section 4 discusses the simulation results and discussions of the proposed model, and Section 5 mentions the conclusion of the overall work.

2 Related works

Some of the existing works carried out on the feature selection process are discussed as follows,

Sathiyabhama et al. [22] introduced a feature selection technique in breast cancer detection using the grey wolf optimizer (GWO) algorithm. The large dimensionality of the feature set is reduced by optimally selecting the required features according to the

GWO concept. This GWO is inspired by the social hierarchy and hunting behaviour of grey wolves. This existing work hybridizes the GWO algorithm with a Rough set (RS) to identify the essential features from the given images. The developed algorithm reduces the irrelevant features from the image and ensures a smooth classification process. The simulation analysis proves that the established algorithm provides better feature selection and helps to attain improved accuracy performance in classification.

Shanthi et al. [23] developed a modified stochastic diffusion search (SDS) approach for optimal feature selection in lung cancer prediction. This approach improves the classification process by obtaining important features that effectively identify different kinds of classes. The SDS algorithm selects the most needed features, and NB and decision tree approaches classify lung cancer. The large size of the features is minimized optimally and gives better classification results. This study reveals that the developed feature selection algorithm is highly suitable for large data quality. The experimental analysis shows that the developed model gives high accuracy in lung cancer prediction.

Asgarnezhad, Razieh et al. [24] established a multi-objective grey wolf (MOGW) algorithm based on feature selection in text prediction. In this study, the text is classified by using the NN classification. KNN chooses the most significant words, and NB approaches via an optimization algorithm. Three different datasets, like TS2, PMD and TS3, are used for the simulation analysis. The developed algorithm optimally chooses essential features and improves the NN classifier's performance. This existing work attains an accuracy level of 95.75%, and the results are compared with several optimization techniques. This multi-objective algorithm is highly effective for the feature selection process in text classification.

Ghosh et al. [25] developed an ant colony optimization (ACO) for feature selection. Here, the wrapper filter is hybridized with the ACO algorithm in which the reduced subset of features is selected. Also, this optimization algorithm reduces the computational complexity using a wrapper approach. This feature selection algorithm is performed in a multi-objective way, and the simulation validation is done using two varied datasets. The objects are classified using KNN and MLP classifiers. The experimental analysis shows that the developed approach is computationally inexpensive and produces excellent classification results by eliminating irrelevant feature sets.

El-Kenawy et al. [26] introduced a hybrid feature selection concept using GWO and particle swarm optimization (PSO). Hybridization is done by attaining the balance between exploration and exploitation. The PSO algorithm is utilized to promote the diversity of the population and enhance production efficiency. The quality is accessed using 17 UCI datasets based on machine learning, which evaluates the optimization algorithm's consistency and ensures the developed solution is stable and reliable. These developed hybrid algorithms are meta-heuristics algorithms, which choose better features and improve the system performance by minimizing the complexity.

Pathak et al. [27] have utilized levy flight based GWO for the feature selection process in image steganalysis application. In order to reduce the irrelevant features from the given inputs, the developed study adopted a new optimization method. This algorithm assisted in choosing the most important features for attaining better steganalysis outcomes. The feature extraction was performed here using AlexNet, Convolutional neural network (CNN) and SPAM. Due to the selection of the most suitable features, the classification accuracy was enhanced, and it exhibited that the feature selection process is necessary for mitigating the large dimensionality issues.

Singh et al. [28] developed a collaborative feature optimization for glaucoma diagnosis using retinal fundus images. The proposed approach was a two-layer approach working on the basis of PSO, Binary Cuckoo Search (BCS) and Bat algorithm. These algorithms

are now said to be single- and double-layer approaches used to combine features. These features are used in the classification process with high accuracy. In the evaluation phase, the developed model used the ORIGA and REFUGE datasets and provided output with 98.95% accuracy. Among the advances was that the model reflected the problem of overfitting due to the combination of uneven layers.

Munish Khanna et al. [29] developed a methodology for human disease prediction based on machine learning approaches. The constructed model used ant lion-based optimization for feature selection, here four classifiers were used for prediction. Three public data sets and one private data set are available to evaluate the model. Five measures were used to evaluate the performance of the proposed model. The result is significantly improved by this change. The method enabled a 50 percent reduction in the original functionality without sacrificing accuracy or performance. For the heart disease dataset: 79.99% for the diabetes dataset, 98.52% for the diabetic retinopathy dataset and 97.18% for the skin cancer dataset to achieve maximum accuracy.

Singh et al. [30] developed three feature selection strategies based on Bacterial Foraging (BFOA), Emperor Penguin (EPO) and a hybrid (hBFEPO) linking BFOA and EPO. For further ML tasks, in addition to the classification of breast cancer, the basic methods for feature selection were examined. For the first time, a hybrid of the two was used. The COVID-19 data set served as the first testing ground for these tactics. These algorithms will be tested on the WDBC breast cancer dataset after producing good results. The model achieved an accuracy of 98.49% in the evaluation phase.

Singh et al. [31] projected a metaheuristic algorithm that worked to replicate the emperor penguin's activities. This proposed algorithm was called emperor penguin optimization, and another optimization combined with this algorithm was bacterial foraging optimization. Here, 36 features were extracted from the retinal fundus images. The proposed approach was used in selecting features with features that significantly improved the accuracy of classification. Six machine learning classifiers classify based on a smaller subset of features provided by these three optimization techniques. The hybrid optimization technique combined with random forest achieved the highest accuracy of up to 0.95410. However, the model is only used for a small subset of functions.

Khan et al. [32] presented a hybrid optimization for efficient feature selection. In this work, we explain a typical feature selection problem to reduce the number of roles and responsibilities while increasing accuracy. From the machine learning repository, another classification dataset and feature selection technique for SMA was tested using GWO. The feature selection for the datasets in the UCI repository includes bat optimization, slime mold optimization, cuckoo search optimization, particle swarm optimization, whale optimization, and grey wolf optimization. Here, the algorithms were evaluated on limited datasets to show the performance of feature selection.

2.1 Problem statement

Feature selection in image processing applications plays a major role in data mining and machine learning. The high number of features in the presented data involves irrelevant and noisy information. There is a higher chance of degrading the classification accuracy and performance. Thus, various feature selection techniques are developed in advanced information technologies. However, many existing approaches face critical challenges regarding appropriate feature selection due to the high increase in search space. Also, some traditional search techniques have the disadvantage of expensive computation, which falls in

local optima and nesting effect. The existing approaches failed to process large data dimensionality, and the previous techniques cannot provide better classification results. In addition, the existing optimization algorithms easily fall into local optima, affecting the system performance. The performance of the search algorithm is enhanced by properly balancing the exploration and exploitation process. Thus, the proposed work develops a hybrid optimization algorithm for an appropriate feature selection.

3 Proposed methodology

This section describes feature selection issues with the KNN classifier, and a hybrid optimization algorithm is introduced to resolve the issues in the feature selection process. Selecting an optimal feature set from a given dataset is a critical problem in machine learning. In recent decades, various techniques have been developed to solve the issues in the feature selection process. Nowadays, meta-heuristic algorithms are becoming more popular in optimal feature selection. Many of the advanced optimization algorithms provide enhanced performance by solving issues in feature selection. The proposed work hybridizes an SMA with BGWO to resolve the feature selection issues with improved classification accuracy. Initially, the input data are pre-processed, where data cleaning is done to clean the given input data. Then, the feature selection stage is performed based on the hybrid optimization techniques. After completing the feature selection phase, the classification is enabled with the aid of the KNN method. Figure 1 represents the block diagram of the proposed work.

3.1 Pre-processing

Pre-processing is a significant stage for image processing because it promotes data quality, reduces redundant data, and ignores the noises available in the provided input high dimensional dataset. The data cleaning process is performed in the pre-processing stage in the proposed work. The process of data cleaning is to identify and remove inaccurate records from a given dataset. Also, it finds incorrect, missing values,

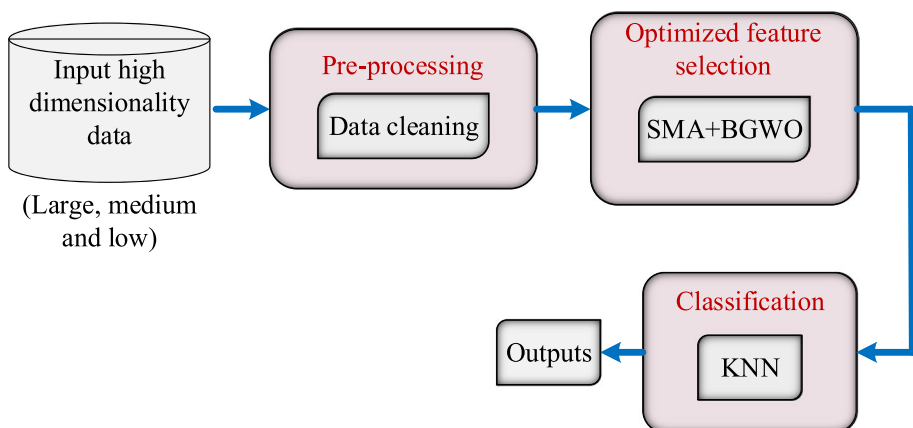


Fig. 1 Proposed block diagram

incomplete and irrelevant data parts and performs modifying, replacing or eliminating the noisy data. In this pre-processing stage, the data is cleaned, thereby enhancing the quality of input data. The pre-processed data are fed to the input of the feature selection techniques.

3.2 Description of different feature selection methodologies

In order to select the optimal features, the proposed work implements eight different types of optimization techniques. From the eight techniques, the best algorithms for the feature selection process are chosen in the proposed work. The best algorithm depends on performance metrics like mean, best, worst, standard deviation and computational time. The algorithms include Ant Colony Optimization (ACO), Bat optimization algorithm (BAT), Cuckoo Search Optimization (CSO), Firefly Optimization Algorithm (FFA), Whale Optimization Algorithm (WOA), Particle Swarm Optimization (PSO), Grey Wolf Optimization (GWO) and Slime Mould Algorithm (SMA) are used for analyzed for attaining better optimization algorithm.

3.2.1 Ant colony optimization (ACO)

The ACO algorithm is one of the nature inspired meta-heuristic approaches to provide resolution of hard optimization problems. The utilization of ACO to identify the most effective subset of features that optimizes the performance of a classification model is possible. When dealing with high-dimensional datasets, where selecting relevant characteristics can improve classification accuracy and reduce computational costs, this is especially advantageous. The proposed work utilizes the ACO approach for feature selection purposes. Each ant is considered a feature, and the features are selected based on the selection probability. Initially, each of the features has a similar probability of selection, represented as follows [25],

$$Q_p(i) = \frac{[\phi(i)]^\alpha [\xi(i)]^\beta}{\sum_{k \in M_i^p} [\phi(k)]^\alpha [\xi(k)]^\beta} \quad (1)$$

where, $\xi(i)$ =feature's document frequency and it mentioned as the number of documents presented in the training set. It involves i^{th} feature and mentions heuristic information presented to the ants. The pheromone trial value is signified as $\phi(i)$, and the two parameters

Table 1 ACO parameter values

Evaporation rate e	0.5
Pheromone factor α	1
Visibility factor β	2
Random variable q	8
Maximum iteration	100
Features	Ants

are given as α and β , which finds the corresponding effect of heuristic and pheromone information. Table 1 shows the range of parameters in the ACO approach.

Based on the global update rule, the pheromone trail is updated, which is mentioned as,

$$\phi(i) = \lambda\phi(i) + \sum_{p=1}^m \Delta\phi_p(i) \quad (2)$$

where, λ mentions the parameter of pheromone evaporation, which degrades the pheromone trails, m be the number of features. The particular number of pheromones is given as $\Delta\phi_p(i)$ in which all the features are stored on the trail and is expressed as,

$$\Delta\phi_p(i) = \begin{cases} 2 U_p F_p & \text{if feature } i \text{ is utilized by search agent } p \\ U_p F_p & \text{if feature } i \text{ is utilized by any search agent } p \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where, F_p be the F-measure value of p^{th} search agent's subset of features and the unit of the pheromone value is mentioned as U_p . The maximized F-measure of the search agent's chosen subset, more pheromone is stored on the features utilized in the subset, and such features are chosen in the upcoming iteration. Based on this, the features are selected by using the ACO algorithm. Moreover, this algorithm has a limited parallelization problem and is computationally complex.

3.2.2 Bat optimization algorithm (BAT)

The BAT approach is one of the meta-heuristic algorithms inspired by the behavior of bats. BAT helps to identify and remove noisy or irrelevant features that improve the signal-to-noise ratio in the data and result in more robust models. This algorithm can be customized to optimize feature subsets based on specific goals or constraints, such as to maximize accuracy, minimize the number of features, or achieve a balance between precision and recall. Initially, the bats utilize echolocation to sense food distance and find the barriers in their specific characteristics. Initially, the features population are initialized, and the parameters are also initialized. The position and velocity of the features are updated for each iteration. When optimal features are selected from a present best solution, a new solution for all search agents is created utilizing a local random walk, which is represented as [33],

$$z_{\text{new}} = z_{\text{old}} + \delta X^t \quad (4)$$

where, δ mentions the random number ranges from [-1, 1], the average attributes of each search agent are mentioned as $X^t = \langle X_i^t \rangle$. The penalty function is introduced for

Table 2 BAT parameter values

Loudness X	0.5
Pulse rate	0.5
Frequency minimum Q_{\min}	0
Frequency maximum Q_{\max}	2
Maximum iteration	100

non-linear equality and inequality constraints. Table 2 shows the range of parameters in the BAT approach.

Based on the penalty function, the constrained issue is converted into an unconstrained issue, which is given as,

$$\rho(z, \eta_i, l_j) = f(z) + \sum_{i=1}^N \eta_i \varpi_i^2(z) + \sum_{j=1}^N l_j \xi_j^2(z) \tag{5}$$

From the above equation, the constraints are managed using the penalty function. The features are ranked and analyzed as the best features in every iteration. The features are selected using the BAT approach depending on the above equations. This algorithm is intended to be sensitive to multiple parameters, and the parameters used may have a lack of intuitive interpretation.

3.2.3 Cuckoo search optimization (CSO)

The CSO approach is one of the algorithms utilized for solving optimization issues. Feature selection using CS helps reduce dimensionality by identifying a subset of the most informative features. Reducing the number of features helps simplify models, shorten training times, and improve the generalization ability of the model. It is also a meta-heuristic algorithm and depends upon the levy flight’s cuckoo species and random walks. This algorithm describes the behavior of cuckoo birds and their strategy of laying an egg in another bird’s nest. The CSO algorithm process is carried out on cuckoo breeding behavior, levy flight, and cuckoo search. The breeding behavior presents three kinds of brood parasitism: cooperative breeding, intraspecific brood parasitism, and nest takeover. In the exploration process, the group with the better quality features are chosen for the next process. At each iteration, the position of optimal features gets updated. When generating a new optimal feature set, a levy flight is enabled and is expressed as [34],

$$z_i^{(t+1)} = z_i^{(t)} + \psi \oplus Levy(\delta) \tag{6}$$

where, ψ mentions the scales of interest problem, $\psi > 0$, the term \oplus represents the entry-wise multiplication. The values of parameters that are fixed in the CSO approach are given in Table 3.

The levy flights perform a random walk, and the random steps are obtained from a Levy distribution for big steps, which is given as,

$$Levy\ v = t^{-\alpha}, \quad (1 < \alpha \leq 3) \tag{7}$$

In this, each step of search agents significantly forms a process of the random walk, which satisfies a distribution of the power-law step length. Based on this, the

Table 3 CSO parameter values

Parameter Q_a	0.25
β	3/2
Maximum iteration	100

Table 4 FFA parameters values

Parameter β	0.20 minimal value
Parameter α	0.5
Absorption coefficient η	1
Maximum iteration	100

Table 5 WOA parameter values

Random number q	0.2
Maximum iteration	100

search agent can explore the optimal feature set in the CSO approach. In addition, this approach also has concerns about the sensitivity of the parameters used for selection and the adaptability to dynamic environments is limited.

3.2.4 Firefly optimization algorithm (FFA)

FFA is a stochastic global optimization approach in which the process is based on the behavior of fireflies. In general, selecting a subset of features using FFA can result in more interpretable models. Understanding the relevance of individual features provides insights into the factors that contribute to the model predictions. FFA can also help identify and eliminate redundant features in a data set. Numerous firefly species are available, and this firefly generates the least and most harmonious flash lighting. The firefly flash lighting has three functions: attracting the mate partners, attracting the potential prey, and affording a warning scheme. Table 4 shows the parameter ranges in FFA.

The search agent moves its position to search for the optimal features.

$$y_i = y_i + \alpha_0 e^{-\eta r_{ij}^2} (y_j - y_i) + \beta(\text{rand} - 0.5) \quad (8)$$

where, $\alpha_0 e^{-\eta r_{ij}^2} (y_j - y_i)$ be the feature characteristics, $\beta(\text{rand} - 0.5)$ represents the parameter of randomization. The random number is denoted as *rand* from the uniform distribution ranges from 0 to 1 and β is signified as noise exiting. The distance between the search agent and features is defined as,

$$r_{ij} = \|y_i - y_j\| \quad (9)$$

where, y_i mentions the position of i^{th} search agent. The optimal solution is attained depending on the fitness function. Thus, the optimal features are selected using the FFA approach [35]. This performance of the algorithm depends on the initialization of the feature population.

3.2.5 Whale optimization algorithm (WOA):

WOA is a nature-inspired meta-heuristic based algorithm which imitates the behavior of whales. WOA can help select a subset of features that improve the generalization ability of a machine learning model. By focusing on the most relevant features, the model is less

likely to overfit the noise in the training data. This approach is motivated by the strategy of bubble-net hunting, and the mathematical model is expressed as three processes: encircling prey, bubble-net attacking mechanism, and prey searching. Table 5 illustrates the parameter values that are fixed in WOA.

Encircling prey The search agent can identify the position of features and encircle them. After determining the optimal search agent, the remaining search agents attempt to update their current position towards the optimal search agent [16]. It is defined as,

$$\vec{E} = \left| \vec{F} \cdot \vec{A}^*(t) - \vec{A}(t) \right| \tag{10}$$

$$\vec{A}(t + 1) = \vec{A}^*(t) - \vec{X} \cdot \vec{E} \tag{11}$$

where, t mentions the present iteration, the coefficient vectors are represented as \vec{X} and \vec{F} , the position vector of the optimal solution is denoted as A^* , the position vector is signified as \vec{A} , the absolute value is termed as $\left| \vec{F} \cdot \vec{A}^*(t) - \vec{A}(t) \right|$, which is a multiplication based on element by element.

Exploitation process-Bubble net attacking scheme The exploitation stage has two processes: shrinking of encircling schemes and spiral updation locations. In this, the position of the search agent is updated as,

$$\vec{A}(t + 1) = \begin{cases} \vec{A}^*(t) - \vec{X} \cdot \vec{E} & \text{if } q < 0.5 \\ \vec{E}' \cdot e^{cd} \cdot \cos(2\pi d) + \vec{A}^*(t) & \text{if } q \geq 0.5 \end{cases} \tag{12}$$

where, the random number is represented as q and ranges from 0 to 1. In this, the search agent can randomly explore the features.

Exploration process – Feature search The search agent can search the features and update their position for each iteration. The process is continued until the optimal set of features is obtained. The mathematical model of the exploration process is given as,

$$\vec{E} = \left| \vec{F} \cdot \vec{A}_{rand} - \vec{A} \right| \tag{13}$$

$$\vec{A}(t + 1) = \vec{A}_{rand} - \vec{X} \cdot \vec{E} \tag{14}$$

where, \vec{A}_{rand} is a position vector in a random manner, which is selected from the present feature set? Adjust the search agent who goes beyond the searching space, and the fitness function is evaluated. Based on this, the features are selected optimally for the classification. Except for selection performance, the algorithm has scalability concerns.

3.2.6 Particle swarm optimization (PSO)

PSO is one of the stochastic optimization approaches inspired by the behaviour of swarms [26]. The utilization of PSO to identify the optimal subset of features that optimizes the performance of a classification model is possible. This is especially beneficial when

Table 6 PSO parameter ranges

V_{\max}	6
Upper bound vector p_{\max}	0.9
Lower bound vector p_{\min}	0.2
Trust parameter a_1	2
Trust parameter a_2	2
Maximum iteration	100

working with high-dimensional datasets, as the selection of pertinent features can enhance classification accuracy and mitigate overfitting. Based on the swarm, the PSO approach performs the searching process. In each iteration, the information about every feature set is integrated to manage the velocity of all dimensions. It is utilized to evaluate the current position of the feature. Table 6 illustrates the parameter ranges in the PSO algorithm.

The position of each feature is updated as,

$$k_{i,t+1}^l = \begin{cases} z_{i,t+1}^l, & \text{if } p(A_{i,t+1}) < p(K_{i,t}) \\ k_{i,t}^l, & \text{otherwise} \end{cases} \quad (15)$$

To update the velocity, the inertia weight was developed, and the current updation formula of velocity is expressed as [26],

$$y_{i,t+1}^l = w * y_{i,t}^l + a_1 * rand * (k_{i,t}^l - z_{i,t}^l) + a_2 * rand * (k_{h,t}^l - z_{i,t}^l) \quad (16)$$

A variant is introduced in the PSO approach with a factor ξ by determining the convergence behaviour. It assures convergence and enhances the rate of convergence, and hence, the velocity is updated as,

$$y_{i,t+1}^l = \xi (y_{i,t}^l + \rho_1 * rand * (k_{i,t}^l - z_{i,t}^l) + \rho_2 * rand * (k_{h,t}^l - z_{i,t}^l)) \quad (17)$$

When the optimal solution is attained, the process is stopped, or it updates the fitness function and continues until a better solution is attained. In this process, the optimal features are selected using the PSO approach. By analyzing the optimization above algorithms based on the performance metrics like mean, best, worst, standard deviation and computational time, the SMA approach is superior to the other algorithms. The second-best technique is GWO, the third is BAT, and the fourth is WOA. Thus, the proposed work uses SMA and GWO for the feature selection. The SMA approach is integrated with GWO to attain an optimal feature set. The reason for using multiple optimizations for feature selection from the features obtained from the sampled data set is to improve the robustness, adaptability and effectiveness of the feature selection process across different data sets. Here, every optimization algorithm has an error in the selection process. A novel hybrid approach is implemented to address drawbacks such as restricted parallelization, limited scalability, parameter sensitivity, and initialization issues.

3.2.7 Slime mould algorithm (SMA)

The slime mould is generally represented as *Physarum polycephalum* and was initially categorized as a fungus. The slime mould is a kind of eukaryote mainly located in humid and cool places. The major nutrition stage of slime mould is Plasmodium. The SMA approach

Table 7 SMA parameter values

Proportional number x	0.03
Random number q	0.5
Maximum iteration	100
W	2

is the same as the swarm-based optimization techniques. Each individual would expand over the process and be managed towards the global optimum in every iteration. The SMA optimization approach is categorized into three phases: initialization, searching and exploiting. Table 7 shows the parameter values in SMA.

Initialization Each of the features from the given data is uniformly and randomly initialized in the overall domain and is mentioned as [36],

$$z_i = R_1(ub - lb) + lb \tag{18}$$

where R_1 mentions the random function in the Gaussian distribution. Each value of features should be initialized in this stage, such as total feature size, the maximum number of iterations max *Iter* etc.

Exploration and exploitation Each feature’s position is updated in every iteration and managed to the global optimum during the exploration and exploitation process. It is represented as,

$$z_i(t + 1) = \begin{cases} R_2.(ub - lb) + lb & R_2 < x \\ z_a + f_a \cdot [Y \cdot z_M(t) - z_N(t)] & R_3 < q \\ f_b \cdot z_i(t) & q \leq R_3 \leq 1 \end{cases} \tag{19}$$

where, $z_i(t)$ and $z_i(t + 1)$ mentions the position of i^{th} feature in the present iteration t and future iteration $t + 1$. The random value in the Gaussian distribution is signified as R_2 , the proportional number is depicted as x , which selects some of the features in a random manner to restart the initialization process, $x = 0.03$ as defaults. In the present iteration, two randomly chosen features are represented as $z_M(t)$ and $z_N(t)$, the two random values in the uniform distribution are mentioned as f_a and f_b with the range of $[-k, k]$ and $[-l, l]$ correspondingly. In this, k and l are represented as the two variables corresponding to the number of iterations and the maximum amount of iteration time is given as,

$$k = k \tanh \left(1 - \frac{t}{\max Iter} \right) \tag{20}$$

$$l = 1 - \frac{t}{\max Iter} \tag{21}$$

Here q is also a proportional number, which limits the option of branches. It corresponds to the global optimal fitness function OF,

$$q = \tanh |P_i - OF| \tag{22}$$

The weight parameter is evaluated as,

$$\omega_{s_i(i)} = \begin{cases} 1 + R_4 \cdot \log \left(1 + \frac{B_f - P_i}{B_f - W_f} \right) & \text{condition} \\ 1 - R_4 \cdot \log \left(1 + \frac{B_f - P_i}{B_f - W_f} \right) & \text{others} \end{cases} \quad (23)$$

where, B_f represents the best fitness value and W_f mentions the worst fitness values among the total fitness values P_i , here $i = 1, 2, \dots, m$. The fitness values of each feature are sorted by the following Eq. (24).

$$p_i = \text{Sort}(P) \quad (24)$$

Here, the SMA algorithm has the advantage of showing both exploratory and exploitative behavior in its foraging strategies. An algorithm inspired by these behaviors could balance exploring a wide range of features and exploiting promising subsets for feature selection. This algorithm is adaptable to any noisy environment and also provides a parallel process of feature selection.

3.2.8 Grey Wolf Optimization

The hunting behaviour of grey wolves inspires the GWO approach. The hunting techniques of grey wolves are tracking the prey, encircling the prey and attacking the prey. Prey is considered the feature in the proposed work, and the wolves are assumed to be a search agent. Table 8 depicts the parameters in GWO.

Encircling the features In the encircling behaviour, the search agents encircle the needed features and are given as [22],

$$\vec{H} = \left| \vec{G} \cdot \vec{Z}_q(t) - \vec{Z}(t) \right| \quad (25)$$

$$\vec{Z}(t+1) = \vec{Z}_q(t) - \vec{U} \cdot \vec{H} \quad (26)$$

where, \vec{U} and \vec{G} are considered as the coefficient vectors, t denotes the number of iterations, \vec{Z}_q be the vector of feature position, the vector of search agent is signified as \vec{Z} and \vec{H} mentions the evaluated vector, which is utilized to represent a search agent's new position. The vectors \vec{U} and \vec{G} are evaluated as,

$$\vec{U} = 2\vec{k} \cdot R_1 - \vec{k} \quad (27)$$

Table 8 GWO parameter values

Random numbers R1 and R2	0.5 and 0.3
Controlling parameter α	0.2
Maximum iteration	100

$$\vec{G} = 2 \cdot \vec{R}_2 \quad (28)$$

where \vec{k} mentions the vector is set to minimize linearly from two to zero among the interactions, the random number ranges from [0, 1] is represented as \vec{R}_1 and \vec{R}_2 . Based on the position of features, the search agent can change its position. The position of search agent is updated as,

$$H_i = \frac{H_\alpha + H_\beta + H_\chi}{3} \quad (29)$$

The hunting process of search agents is managed by parameters like α , and β and χ . H_α Mentions the first optimal search agent, H_β represents the second optimal search agent and H_χ denotes the third optimal search agent.

Attacking the features The search agents complete the search process by attacking the features until they stop moving. The attacking process is modeled by minimizing the value of \vec{k} in varied iterations. If the fluctuation rate of \vec{k} reduces, then the vector \vec{U} also reduces. Based on the position of the parameters α , and β and χ , the search agent updates its position.

Exploration process The search agents often explore for the features based on the positions of parameters α, β and χ . The search agents diverge from each other and associate with attacking the features to search for the feature's position. The GWO approach has a component (\vec{G}) which helps the optimization algorithm to generate new solutions. This component provides random weights for each of the features, and it assists in maintaining the search agent away from local optima issues. The weights of the GWO approach are given as follows,

$$\psi = \frac{1}{2} v \tan(t) \quad (30)$$

$$\theta = \frac{2}{\pi} v \cos \frac{1}{3} \cdot v \tan(t) \quad (31)$$

$$w_1 = \cos \theta, w_2 = \frac{1}{2} \sin \theta \cdot \cos \psi, w_3 = 1 - w_1 - w_2 \quad (32)$$

In addition, the GWO algorithm has high interoperability and is inherently parallelizable, making it suitable for parallel computing environments. This property can be advantageous when processing computationally intensive optimization problems.

3.3 Feature selection using the proposed hybridized SMA + BGWO model

The feature selection process reduces the large dimensionality of the features from the pre-processed data. This feature selection phase helps improve overall system performance. In addition, the feature selection phase shortens the training time of classification and improves accuracy. Different algorithms are analyzed to select the better optimization algorithm suitable for effective classification. First, the algorithms ACO,

BAT, CSO, FFA, WOA, PSO, GWO and SMA are analyzed. Among these algorithms, SMA and GWO algorithms are selected as better optimizers for feature selection due to their high interoperability, scalability and inherently parallelizable parameters compared to the other optimization algorithms. Therefore, the proposed work utilizes a hybrid form of SMA and GWO based optimization for the feature selection process. In order to make the system more effective and achieve better convergence, the GWO algorithm is changed to BGWO in addition to the SMA algorithm. The fitness function of hybrid algorithms is represented as follows.

$$\text{Fitness function} = \delta e_r(C) + \eta \frac{|S|}{|N|} \quad (33)$$

where, $e_r(C)$ mentions the classification error rate, the total volume of features is represented as N . The chosen subset of features is denoted as S , δ and η mentions the parameters related to the accuracy of the classification process and the minimization of features. Depending on the fitness function, the features are selected using hybrid algorithms. The combination of SWA and BGWO aims to accelerate convergence by integrating effective mechanisms from each algorithm. This can lead to faster convergence towards optimal solutions, especially for high-dimensional and complex optimization problems.

3.3.1 Hybridization process

The SMA algorithm is hybridized with BGWO to obtain a better feature selection outcome. In this, the best feature z_a is replaced with the average of α, β and χ , the hybridization process is updated as,

$$z_i(t+1) = \begin{cases} R_2 \cdot (ub - lb) + lb & R_2 < x \\ w_1 H_\alpha + w_2 H\beta + w_3 H\chi + f_a \cdot [Y \cdot z_M(t) - z_N(t)] & R_3 < q \\ f_b \cdot z_i(t) & q \leq R_3 \leq 1 \end{cases} \quad (34)$$

To perform fine-tuning of the hybrid algorithm, the BGWO approach is introduced in the proposed work. This BGWO approach is combined with SMA to improve the performance of the optimization process and also allows to obtain better features. The hybridization process can enable parallel exploration of the solution space by different algorithms. This parallelism can lead to more efficient use of computing resources. In the presence of noise in the objective function, hybridization can provide robustness. Different algorithms respond differently to noise, and the hybrid approach can mitigate the impact of noise on the optimization process.

3.3.2 Binary grey wolf optimization (BGWO- (Sigmoid transfer function (S-shape))

Based on this hybridization process, the features are optimally selected. In the fundamental optimization algorithms, the individual features move in the search space to update their positions to any location in the space, termed continuous space. For the basic feature selection problems, the solutions are restricted to the binary space $\{0, 1\}$ values. To resolve the issues in feature selection, the continuous space must be converted to their related binary solutions $\{0, 1\}$. Therefore, it inspired the introduction

of the latest version based on a sigmoid transfer function (S-shape). The updation of the S-shaped function is mentioned as,

$$v^p = \frac{1}{1 + e^{-y_i^p(t)}} \quad (35)$$

$$Z_i^d = \begin{cases} 1 & \text{if } rand < L(z_i^p(t+1)) \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

By applying the S-shaped function, a free position of the search agents must be converted to their related binary solutions. This conversion is performed in each dimension, representing the probability of changing the elements of position vectors from 0 to 1. Hence, the search agents are moved to a certain binary space. This S-shaped function makes the performance of the hybrid optimization process high. Table 9 gives the pseudocode for the optimized feature selection process.

By analyzing the performance of each feature selection approach, SWA and BGWO are selected as effective approaches. Thus, these two approaches are hybridized in the proposed study to make a robust feature selection. In order to further enhance the feature selection performance in the proposed study, the existing GWO is enhanced by BGWO, and then hybridization is enabled over SWA. Because of utilizing BGWO, the local minima issues are avoided and the convergence rate is enhanced compared to other existing feature selection algorithms. This hybridization in existing work has been evaluated with limited datasets. To show the improved performance in the proposed work, the hybrid algorithm is evaluated with many datasets and for improvement, the sigmoid transfer function is also used for optimization. In general, numerical features offer various advantages when used on high-dimensional datasets. Dimensionality reduction approaches are often better suited to numerical features. Compared to complex categorical properties, numerical features are typically easier to understand. Numerical data tends to scale more smoothly as the data set grows. Large datasets with numeric features are typically easier to process than datasets with categorical features. For feature development, numerical features offer more flexibility. In metaheuristic algorithms, there is the freedom to combine, alter, or interact with current numerical characteristics to build new ones, giving the ability to capture more complex data relationships.

3.4 KNN based classification

The selected optimal features are subjected to the classification stage in which different kinds of data and their categories are accurately classified. The proposed work uses the KNN approach for classification purposes. The KNN classification method is more popular because of its high efficiency and simplicity. This model learns from the training data during the prediction phase, making it computationally efficient during the training process. Furthermore, it is used for binary and multiclass problems, while ANN makes no assumptions about the underlying data distribution. This flexibility makes it suitable for a wide range of data sets, including those with non-linear decision boundaries. This KNN approach generates a decision on categorizing new testing data with the known training data. Basically, for a provided unlabeled time series Z , the rule of the KNN method identifies the K “neighborhood” labeled time series from the training data. Also,

Table 9 Pseudo code of proposed feature selection process

```

Execute SMA algorithm;
Initialize the number of features
Initialize the position of the feature set
While ( $t \leq Max\_iteration$ )
  For  $i = 1$  to  $n$ 
    Initialization of features by equation (18);
    Update the position of features by equation (19);
    Update the fitness value using equation (22);
    Evaluate the weight parameter by equation (23);
    Sort the fitness of each feature by equation (24);
  End for
   $t = t + 1$ ;
End While
  Return Optimal Solution;
Execute GWO algorithm;
  Initialize the total amount of features;
  Initialize the parameters;
  Evaluate the fitness function;
  Update the position of search agent;
  Apply weights for GWO by equation (30), equation
(31) & equation (32);
  Fitness value of hybridization using equation (33);
  Update the hybridization process by equation (34);
Execute BGWO;
  For  $i = 1$  to  $n$ 
    Update S-shape using equation (35) & equation
(36);
  end for
  If the exiting search agent is from the search space
  then
    Compute the fitness function
  end if
Return \ \ Optimal Features

```

it allocates Z to the class which repeatedly belongs in the closest of the k time series. In KNN, the classification process is performed by a maximum vote of the nearest neighbour. The Euclidean distance between the two data samples is evaluated as,

$$Dist(A, B) = \sqrt{\sum_{i=1}^D (A_i - B_i)^2} \quad (37)$$

where, A_i and B_i are the attributes of two different samples. The test data are classified using the following two voting methods such as majority voting and distance weighted voting, which are given as,

$$\text{Majority voting, } m' = \arg \max_f \sum_{(z_i, m_i) \in D_x} \psi(f, m_i) \quad (38)$$

$$\text{Distance – weighted voting, } m' = \arg \max_f \sum_{(z_i, m_i) \in D_x} \omega_i \psi(f, m_i) \quad (39)$$

where, $\psi(f, m_i)$ represents the indicator function, the set of the closest neighbour of the test data is signified as D_x and ω_i mentions the weight in which $\omega_i = 1/d(z', z_i)^2$. KNN implicitly performs feature selection by considering only the relevant features for determining the similarity between data points. Irrelevant features have less impact on the algorithm's decision. By using the KNN approach, the classification is performed with reduced error.

4 Results and discussions

This section evaluates the performance of the proposed techniques using different high dimensional datasets, and the attained results are compared with several existing approaches to compute the efficacy of the proposed methods. Different feature selection algorithms are implemented using seven varied datasets, and the results are compared with metrics like mean, worst, best, computational time and standard deviation. Based on the results, the best feature selection algorithm is chosen for the further process. Also, different classifiers like SVM, NB and KNN are executed for the classification stage. The best classifier is chosen according to some of the metrics from these classifiers. Then, the hybrid optimization for the feature selection process is performed using the best two algorithms, and the chosen best classifier does the classification.

4.1 Dataset description

Different high dimensional datasets are used for the experimental setup in the proposed work. Initially, seven datasets are utilized to analyze the optimal algorithm for feature selection: Banknote, Breast Cancer, Diabetes, Heart, Liver, Alzheimer's and Zoo. These datasets are collected from the UCI machine learning repository.

Banknote authentication dataset In the banknote authentication dataset, the data are gathered from the images acquired to compute an authentication concept of banknotes. This dataset involves 1372 instances, and the number of attributes is 5.

Breast cancer dataset The total number of instances in the breast cancer dataset is 286, and the available amount of attributes is 9. This dataset has multivariate characteristics and contains the attributes information about age, class, size of the tumor, etc.

Diabetes dataset The diabetes dataset involves 20 different attributes and the files of diabetes patients with four different fields per record.

Heart disease dataset The heart disease dataset contains 75 kinds of attributes with 303 instances. This dataset holds the details of heart disease patients, including age, sex, id, etc.

Liver dataset The liver disorder dataset has 345 instances and contains 7 attributes.

Alzheimer's The Alzheimer's dataset is collected from the Kaggle repository with several MRI-based images. The four categories in the Alzheimer's dataset are mild demented, non-demented, moderate demented and very mild demented. Moreover, this dataset is manually collected from several websites.

Zoo dataset The Zoo dataset involves 101 different instances and 17 volumes of attributes. The available information of attributes is milk, airborne, feathers, legs, eggs, etc.

Using such datasets, eight feature selection algorithms and three varied classifiers are executed, and from that, the best techniques are chosen using performance metrics. After that, the selected techniques are executed using a large data dimensionality.

High dimensionality data The proposed work uses the dataset with high dimensionality by categorizing the dataset into large, medium and low. In the large dataset, two datasets are taken in the proposed experimental setup, namely CICDDoS2019 and CICMalDroid2020. The proposed work uses three datasets in the medium dataset: heart dataset, wine quality dataset, and car sales dataset. The low dataset used is the banknote, seed and Prima Indians Diabetes.

CICDDoS2019 The CICDDoS2019 contains 12 classes of attacks and presents TFTP attack data with a ratio of 100 to 1. This dataset is collected from the Kaggle DDOS dataset for several network security-based approaches.

CICMalDroid2020 The CICMalDroid2020 dataset has 17,341 samples gathered from Dec 2017 to Dec 2018. This dataset is classified as Banking malware, Riskware, Adware, Benign and SMS malware.

Wine quality dataset The wine quality dataset has 4898 instances, and it also includes 12 amount of attributes. This dataset has white and red Vinho Verde samples of wine collected from Portugal.

Car sales dataset The car sales dataset is acquired from the Kaggle repository. This dataset has several pieces of information regarding varied types of cars.

Seed dataset The seed dataset is collected from the BSMI laboratory, which holds 20 subjects, 10 females and 10 males. This dataset is highly utilized in emotion classifications.

Prima Indians Diabetes The Prima Indians Diabetes dataset is gathered from the kidney and digestive diseases of the diabetes institute. This dataset presents the patient 21 years old of a female with diabetes.

4.2 Performance metrics

The performance metrics such as accuracy, precision, recall, F-measure, time, MSE and MAE are utilized to evaluate the proposed classification scheme. The metrics like best, worst, mean, standard deviation and computational time are utilized to compute

the efficacy of the optimized feature selection techniques. The efficacy of the proposed work is measured by comparing the attained results with various existing techniques.

4.2.1 Metrics for feature selection

The following performance metrics are measured to analyze the effectiveness of feature selection techniques.

Mean fitness The mean fitness represents the average value of the attained fitness function. When the proposed algorithm is processed at N times, then the mean fitness is evaluated as,

$$\text{mean}_f = \frac{1}{N} \sum_{p=1}^N h_*^p \quad (40)$$

where, h_*^p be the fitness function value.

Best fitness The best fitness mentions the lowest value of the fitness function. When the technique is processed at N times, then the best fitness is computed as,

$$\text{Best}_f = \text{Low}_p^N = 1^{h_*^p} \quad (41)$$

where, h_*^p signifies the best fitness value.

Worst fitness The worst fitness mentions the highest values of the fitness function. When the feature selection technique is processed at N times, the worst fitness is evaluated as,

$$\text{worst fitness} = \text{High}_{p=1}^N h_*^p \quad (42)$$

where, h_*^p represents the worst fitness value.

Standard deviation The standard deviation is the metric to measure the stability of each optimization algorithm in the feature selection process. Based on the measure of the fitness function, the standard deviation is evaluated and is given as,

$$\text{Standard deviation} = \sqrt{\frac{1}{M} \sum_{p=1}^M (x - \mu)^2} \quad (43)$$

where, M be the total features, x be the value of data, and the mean value is denoted as μ .

Computational time Computational time is the metric used to evaluate the average time for the computation process in sec. When the feature selection approach is run at N times, then the average computational time is expressed as,

$$\text{Avg. time} = \sum_{p=1}^N \text{Average time}^p \quad (44)$$

where, Average time^p mentions the average processing time.

Table 10 Performance analysis of proposed feature selection techniques

Data Set	FS Algorithm	Selected Features	Best	Worst	Mean	Standard deviation	Computational Time
Data Set 1 (Diabetics) (9)	PSO	6	2.1051	2.478	2.2456	0.0507	1.835
	ACO	3	1.2146	1.5481	1.3178	0.0101	3.1875
	FFA	6	0.4177	0.9751	0.6843	0.0465	10.6545
	CS	5	0.1348	0.2732	0.2518	0.0753	1.378
	WOA	6	0.0254	0.5789	0.0387	0.0047	0.9321
	BAT	6	0.0234	0.0478	0.0347	0.0034	0.3384
	GWO	4	0.0189	0.0578	0.0375	0.0029	1.2542
	SMA	3	0.0167	0.0476	0.0333	0.0037	0.0635
Data Set 2 (Alzheimer) (12)	PSO	5	1.2159	2.4318	1.854	0.0638	3.313
	ACO	6	0.3375	0.5575	0.4258	0.0554	4.8324
	FFA	6	0.3548	0.5756	0.3678	0.0175	6.3971
	CS	5	0.0428	0.0684	0.0579	0.0067	1.3140
	WOA	6	0.0374	0.0634	0.0475	0.0054	1.3421
	BAT	4	0.0281	0.0598	0.0421	0.0061	0.4354
	GWO	5	0.0183	0.0518	0.0364	0.0031	1.5872
	SMA	4	0.0175	0.0417	0.333	0.0043	0.0640
Data Set 3 (Heart)	PSO	11	0.621	1.0130	0.8791	0.1683	2.0748
	ACO	6	0.4651	0.7407	0.5964	0.1149	0.9204
	FFA	9	0.0896	0.4231	0.1591	0.0871	14.7208
	CS	8	0.0408	0.0816	0.0608	0.0382	0.9058
	WOA	5	0.0176	0.0213	0.0192	0.0020	1.0764
	BAT	7	0.0114	0.0214	0.0158	0.0037	0.6240
	GWO	7	0.075	0.0090	0.008	0.0048	2.7622
	SMA	6	0.0196	0.370	0.333	0.0047	0.1092
Data Set 4 (Bank Note) (4)	PSO	4	0.5213	0.8897	0.7138	0.0617	2.7755
	ACO	4	0.1726	0.7156	0.3891	0.0173	0.6639
	FFA	4	0.0816	0.1421	0.1047	0.0498	3.7891
	CS	4	0.0698	0.0974	0.0752	0.0081	3.6043
	WOA	4	0.0348	0.0408	0.0393	0.0053	1.5596
	BAT	4	0.0274	0.0394	0.0327	0.0049	0.7640
	GWO	3	0.0334	0.0642	0.0514	0.0051	0.0371
	SMA	2	0.0222	0.0333	0.0287	0.0047	0.0300
Data Set 5 (Liver) (11)	PSO	4	1.7760	3.4714	2.8089	0.74608	2.6651
	ACO	4	0.7752	1.2346	0.9500	0.1228	0.3620
	FFA	6	0.1446	0.2754	0.2193	0.0132	20.7099
	CS	3	0.6068	0.7709	0.6879	0.1061	2.0865
	WOA	5	0.2410	0.3038	0.7961	0.0251	1.4791
	BAT	2	0.0237	0.02514	0.0246	0.0051	0.1080
	GWO	3	0.0964	1.1157	0.0996	0.0067	0.3663
	SMA	2	0.0222	0.0667	0.0333	0.0047	0.0680

Table 10 (continued)

Data Set	FS Algorithm	Selected Features	Best	Worst	Mean	Standard deviation	Computational Time
Data Set 6 (Zoo) (19)	PSO	15	0.5132	0.8974	0.6971	0.9816	1.9801
	ACO	14	0.2326	0.3704	0.2917	0.9501	1.6951
	FFA	10	0.0108	0.0135	0.0121	0.0949	5.2112
	CS	11	0.0040	0.0057	0.0048	0.0443	0.3350
	WOA	12	0.0033	0.0039	0.0035	0.0101	1.3210
	BAT	5	0.0038	0.0044	0.0041	0.0091	0.1020
	GWO	8	0.0013	0.0016	0.0014	0.0080	6.3123
	SMA	6	0.0185	0.0556	0.0333	0.0047	0.1500
Data Set 7 (Breast Cancer) (11)	PSO	9	0.7183	0.9515	0.8218	0.7182	3.9136
	ACO	7	0.3226	0.3464	0.3166	0.6081	0.0460
	FFA	7	0.2550	0.2859	0.2676	0.4512	5.6573
	CS	6	0.2291	0.4306	0.3450	0.3178	0.8446
	WOA	5	0.0384	0.0548	0.0427	0.0068	2.7042
	BAT	5	0.0255	0.0454	0.0361	0.0053	0.08134
	GWO	7	0.0093	0.0112	0.0102	0.0018	0.8915
	SMA	4	0.0214	0.0507	0.0333	0.0041	0.0657

4.2.2 Metrics for classification:

The following metrics used to compute the efficiency of the classification approach are discussed below.

Accuracy Accuracy is an essential metric to measure efficiency, and the enhancement rate is the proposed classifier over state-of-the-art techniques. An accuracy metric plays a significant role in determining the performance of the classification process to categorize the given data. The evaluation of accuracy is expressed as,

$$A_c = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \quad (45)$$

Precision Precision is important to evaluate the performance of classification methods; also, it mentions the rate of correctly classified data based on the total amount of true positives. It is acquired as the sum of correctly labelled data from a certain class divided by the entire classified data. It is computed as,

$$P_r = \frac{Tp}{Tp + Fp} \quad (46)$$

Recall Recall is one of the performance metrics to compute the accuracy of the entire model. The total amount of correctly classified data to the overall input data should have

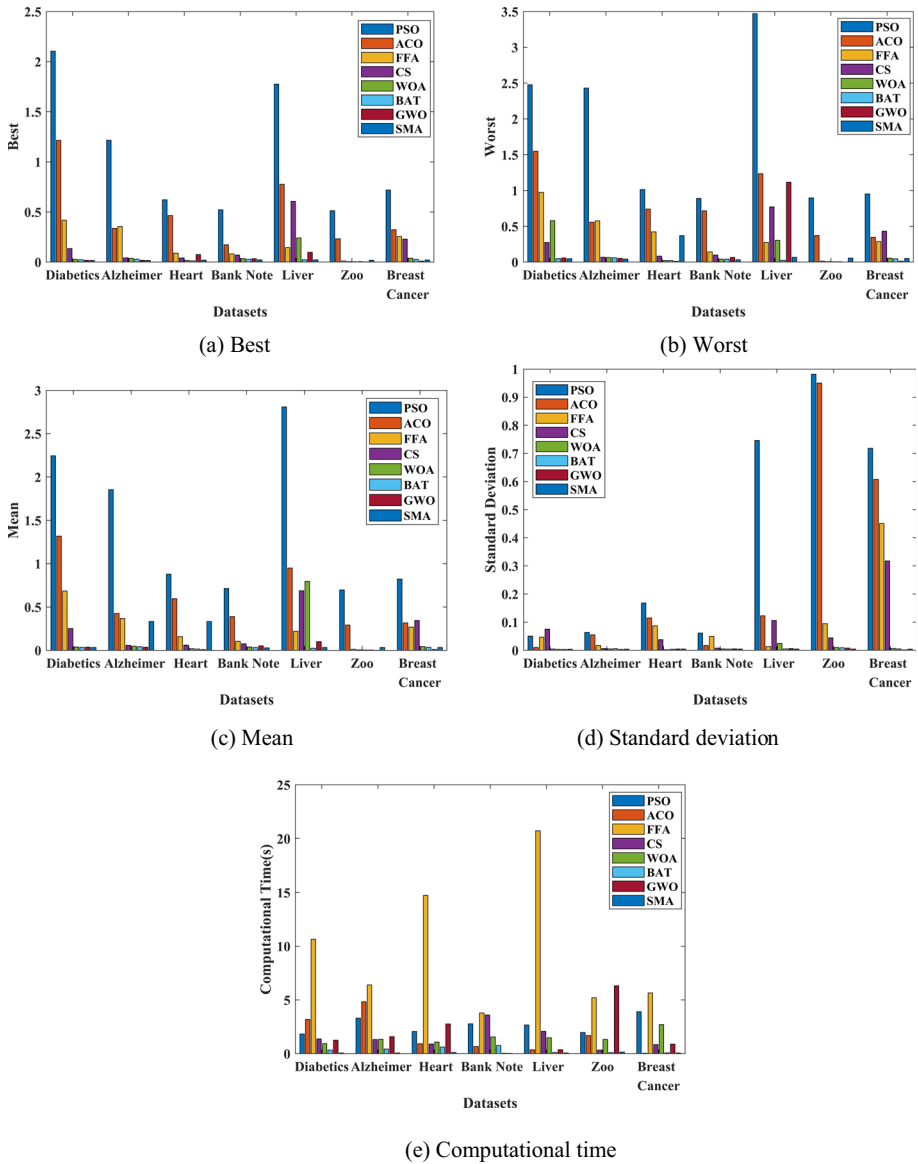


Fig. 2 Performance results of several feature selection techniques using a varied dataset

been classified for a certain label. The recall can identify the error rate in the classification process. It is evaluated as,

Table 11 Performance analysis of proposed hybrid optimization technique with other optimization algorithms

Data Set	FS Algorithm	Selected Features	Best	Worst	Mean	Standard deviation	Computational Time
Data Set 1 (Diabetics) (9)	WOA	6	0.0254	0.5789	0.0387	0.0047	0.9321
	BAT	6	0.0234	0.0478	0.0347	0.0034	0.3384
	GWO	4	0.0189	0.0578	0.0375	0.0029	1.2542
	SMA	3	0.0167	0.0476	0.0333	0.0037	0.0635
	SMA + BGWO	2	0.0145	0.0395	0.0253	0.0027	0.0875
Data Set 2 (Alzheimer) (12)	WOA	6	0.0374	0.0634	0.0475	0.0054	1.3421
	BAT	4	0.0281	0.0598	0.0421	0.0061	0.4354
	GWO	5	0.0183	0.0518	0.0364	0.0031	1.5872
	SMA	4	0.0175	0.0417	0.333	0.0043	0.0640
	SMA + BGWO	3	0.0158	0.0375	0.0278	0.0038	0.0845
Data Set 3 (Heart)	WOA	5	0.0176	0.0213	0.0192	0.0020	1.0764
	BAT	7	0.0114	0.0214	0.0158	0.0037	0.6240
	GWO	7	0.075	0.0090	0.008	0.0048	2.7622
	SMA	6	0.0156	0.0370	0.0233	0.0047	0.1092
	SMA + BGWO	4	0.0145	0.0471	0.0257	0.0037	0.0796
Data Set 4 (Bank Note) (4)	WOA	4	0.0348	0.0408	0.0393	0.0053	1.5596
	BAT	4	0.0274	0.0394	0.0327	0.0049	0.7640
	GWO	3	0.0334	0.0642	0.0514	0.0051	0.0371
	SMA	2	0.0222	0.0333	0.0287	0.0047	0.0300
	SMA + BGWO	2	0.0189	0.0394	0.0256	0.0039	0.0657
Data Set 5 (Liver) (11)	WOA	5	0.2410	0.3038	0.7961	0.0251	1.4791
	BAT	2	0.0237	0.02514	0.0246	0.0051	0.1080
	GWO	3	0.0964	1.1157	0.0996	0.0067	0.3663
	SMA	2	0.0222	0.0667	0.0333	0.0047	0.0680
	SMA + BGWO	3	0.0197	0.0428	0.0256	0.0038	0.0845
Data Set 6 (Zoo) (19)	WOA	12	0.0033	0.0039	0.0035	0.0101	1.3210
	BAT	5	0.0038	0.0044	0.0041	0.0091	0.1020
	GWO	8	0.0013	0.0016	0.0014	0.0080	6.3123
	SMA	6	0.0185	0.0556	0.0333	0.0047	0.1500
	SMA + BGWO	4	0.0125	0.0514	0.0371	0.0035	0.0687
Data Set 7 (Breast Cancer) (11)	WOA	5	0.0384	0.0548	0.0427	0.0068	2.7042
	BAT	5	0.0255	0.0454	0.0361	0.0053	0.08134
	GWO	7	0.0093	0.0112	0.0102	0.0018	0.8915
	SMA	4	0.0214	0.0507	0.0333	0.0041	0.0657
	SMA + BGWO	4	0.0176	0.0645	0.0418	0.0037	0.0482

$$R_e = \frac{Tp}{Tp + Fn} \quad (47)$$

F-measure The F1 score represents the comparative view of precision and recall; the perfect precision and recall are determined by 1 in the F1 score. The F1 score is considered zero when the precision or recall is zero. The F-measure is known as the harmonic mean of precision and recall, which can be formulated as,

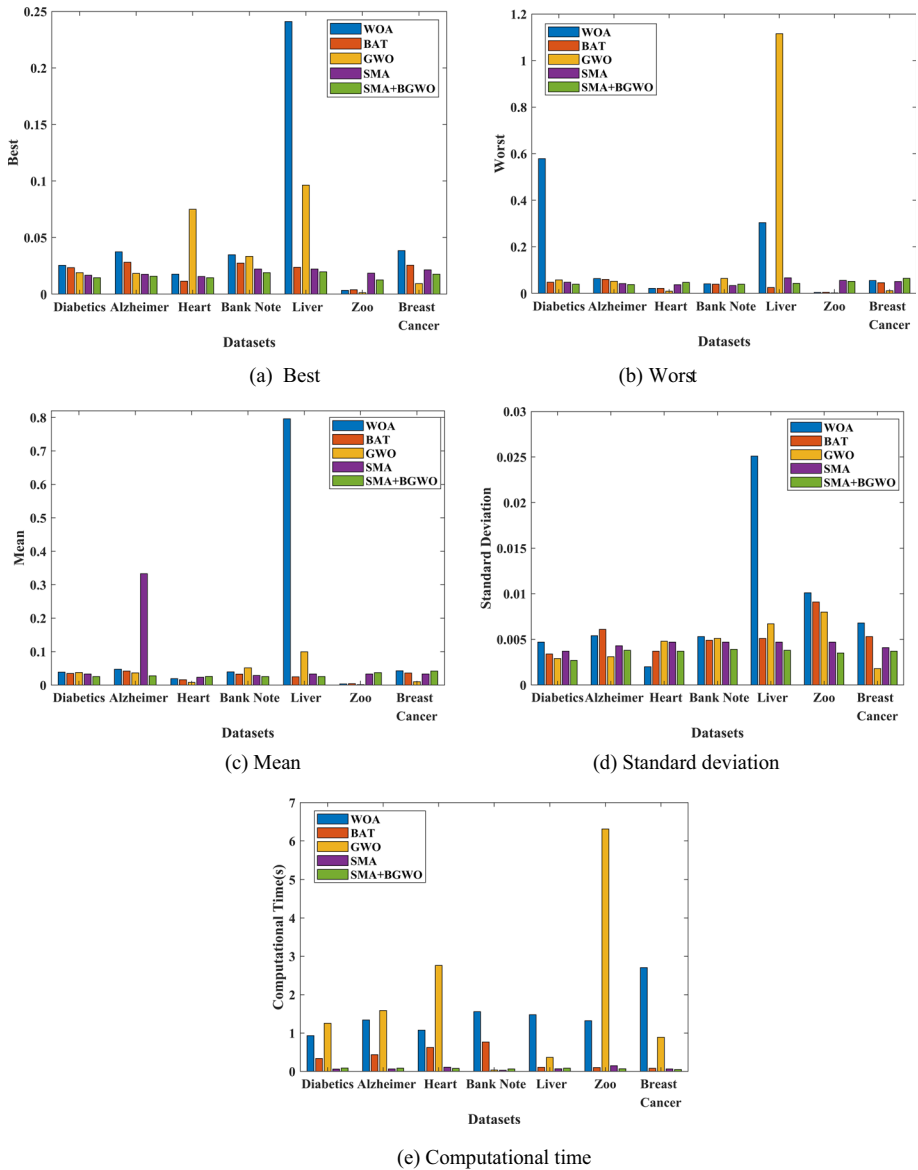


Fig. 3 Performance comparison of the proposed hybrid optimization approach with other optimization techniques

$$F - measure = 2 \times \frac{Precision \times recall}{Precision + recall} \tag{48}$$

RMSE The term RMSE is referred to as Root Mean Square Error, in which the square root of the obtained error is evaluated. The RMSE value is similar to the real units of the target value that are being classified. The RMSE is evaluated as,

Table 12 Performance comparison of proposed and existing hybrid approaches

Data Set	FS Algorithm	Selected Features	Best	Worst	Mean	Standard deviation	Computational Time
Data Set 1 (Diabetics) (9)	BAT-PSO	4	0.0195	0.1971	0.0677	0.0051	0.6741
	ALO-GWO	4	0.0183	0.0678	0.0331	0.0045	0.2471
	GWO-WOA	4	0.0176	0.0478	0.0335	0.0031	0.5213
	SMA-FA	3	0.0151	0.0476	0.0299	0.0028	0.1681
	SMA + BGWO	2	0.0145	0.0395	0.0253	0.0027	0.0875
Data Set 2 (Alzheimer) (12)	BAT-PSO	4	0.0578	0.0761	0.0632	0.0067	2.1231
	ALO-GWO	3	0.0411	0.0782	0.0514	0.0059	1.4912
	GWO-WOA	4	0.0273	0.0398	0.0304	0.0045	1.0102
	SMA-FA	4	0.0269	0.0317	0.0293	0.0041	0.8079
	SMA + BGWO	3	0.0158	0.0375	0.0278	0.0038	0.0845
Data Set 3 (Heart) (15)	BAT-PSO	5	0.0246	0.0375	0.0279	0.0041	1.3124
	ALO-GWO	3	0.0218	0.0289	0.0241	0.0040	0.4147
	GWO-WOA	4	0.0178	0.0210	0.0193	0.0038	1.0245
	SMA-FA	4	0.0168	0.0481	0.0342	0.0041	0.0865
	SMA + BGWO	4	0.0152	0.0372	0.0295	0.0037	0.0796
Data Set 4 (Bank Note) (4)	BAT-PSO	4	0.0781	0.1032	0.0871	0.0072	3.3790
	ALO-GWO	3	0.3174	0.8410	0.6051	0.0058	1.5020
	GWO-WOA	2	0.2274	0.4912	0.3672	0.0049	0.3478
	SMA-FA	2	0.04627	0.0510	0.0488	0.0048	0.6871
	SMA + BGWO	2	0.0189	0.0394	0.0256	0.0039	0.0657
Data Set 5 (Liver) (11)	BAT-PSO	4	0.7112	1.0271	0.8623	0.0091	1.0011
	ALO-GWO	3	0.3671	0.6748	0.4516	0.0075	0.8164
	GWO-WOA	2	0.1147	1.0997	0.7106	0.0071	1.0178
	SMA-FA	2	0.0354	0.0914	0.0674	0.0063	0.5173
	SMA + BGWO	3	0.0197	0.0428	0.0256	0.0038	0.0845
Data Set 6 (Zoo) (19)	BAT-PSO	8	0.1078	0.3784	0.2517	0.0092	0.6807
	ALO-GWO	8	0.7088	2.5641	1.8451	0.1581	1.9710
	GWO-WOA	5	0.6710	1.3179	0.9718	0.0080	0.2713
	SMA-FA	4	0.0773	0.2478	0.1753	0.0053	0.2283
	SMA + BGWO	3	0.0125	0.0514	0.0371	0.0035	0.0687
Data Set 7 (Breast Cancer) (11)	BAT-PSO	6	2.1371	3.518	2.7325	0.1107	1.9700
	ALO-GWO	5	0.9747	1.1028	0.9907	0.8705	1.2157
	GWO-WOA	4	0.4723	0.7875	0.5961	0.0605	0.6517
	SMA-FA	4	0.0578	0.2568	0.1975	0.0108	0.1786
	SMA + BGWO	3	0.0176	0.0645	0.0418	0.0037	0.0482

$$RMSE = \sqrt{\sum_{i=1}^m \frac{(\hat{x}_i - x_i)^2}{m}} \tag{49}$$

where, \hat{x}_i be the predicted value and x_i be the actual value. The total amount of data is represented as m .

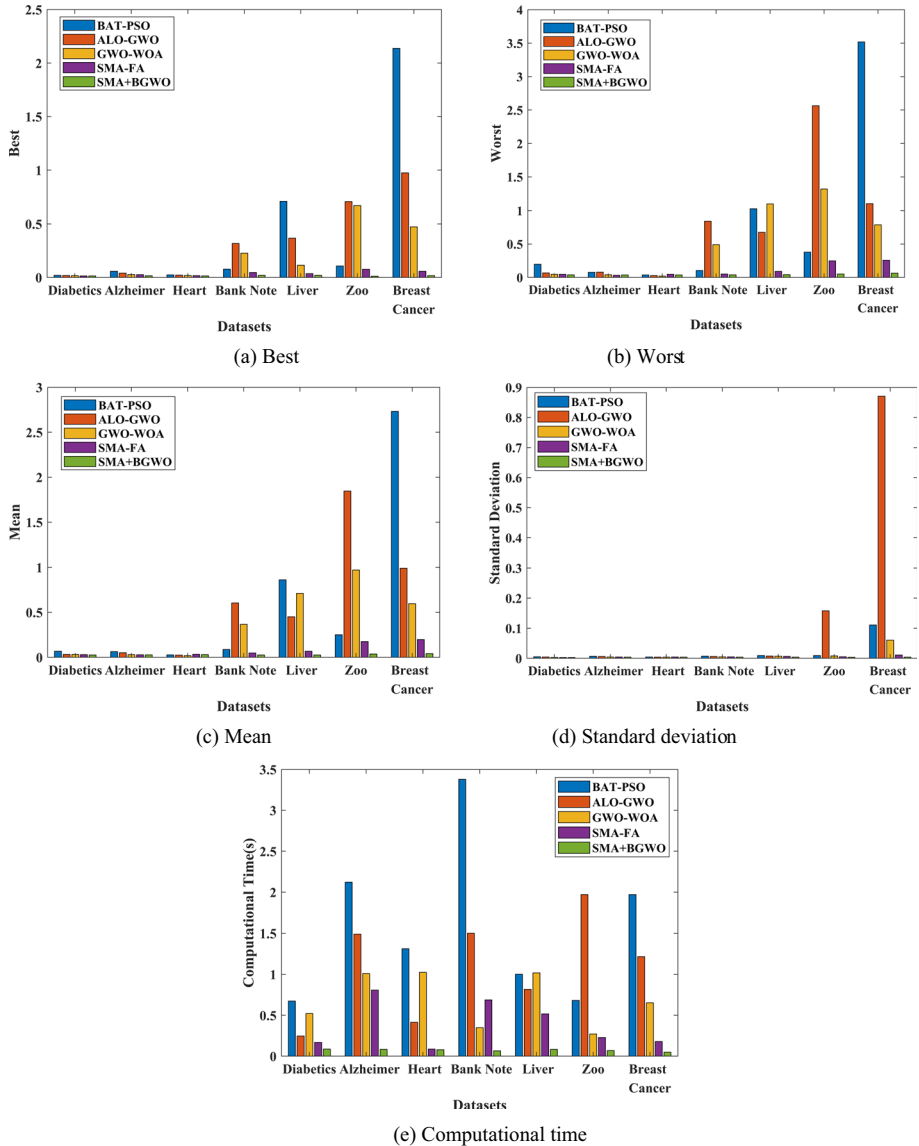


Fig. 4 Comparison analysis of proposed hybrid with existing hybrid techniques using varied datasets

MAE The term MAE stands for mean absolute error, an error measure between the predicted and true values. It is measured as,

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{m} \tag{50}$$

where, x_i denotes the prediction value, y_i represents the true value and m signifies the total amount of data.

Table 13 Classification performance comparison (SMA+GWO)

Data Set	CLASSIFIER	Accuracy	Precision	Recall	F-Measure	Time	RMSE	MAE
Alzheimer's	NB	0.8682	0.9721	0.8345	0.7889	0.0165	5.5321	1.2214
	SVM	0.9797	0.9681	0.9565	0.9614	0.0156	5.5578	0.9637
	KNN	0.9884	0.9761	0.9764	0.9664	0.0123	5.4902	0.8779
Banknote	NB	0.8317	0.8434	0.8569	0.8553	0.0273	0.4048	0.1675
	SVM	0.9879	0.9965	0.9868	0.9842	0.0189	0.0431	0.0127
	KNN	0.9872	0.9986	0.9997	0.9965	0.0148	0.0454	0.0024
Breast Cancer	NB	0.8772	0.8697	0.8668	0.8543	0.1698	0.5828	0.3783
	SVM	0.9464	0.9547	0.9589	0.9136	0.1246	0.6043	0.2783
	KNN	0.9818	0.9768	0.9856	0.9487	0.0538	0.5357	0.2783
Diabetes	NB	0.7773	0.8054	0.8797	0.8451	0.0475	0.4745	0.2251
	SVM	0.6964	0.7768	0.7835	0.7754	0.0314	0.5505	0.3030
	KNN	0.9663	0.9824	0.9013	0.9452	0.0156	0.3879	0.1513
Heart	NB	0.9713	0.9784	0.9513	0.9742	0.0156	0.1483	0.1482
	SVM	0.9748	0.9635	0.9364	0.9651	0.0156	0.1367	0.1174
	KNN	0.9887	0.9812	0.9564	0.9754	0.0024	0.1139	0.0220
Liver	NB	0.8811	0.9181	0.9041	0.8954	0.0781	0.6459	0.4171
	SVM	0.9553	0.92354	0.9116	0.9641	0.0412	0.5904	0.3486
	KNN	0.9623	0.9854	0.9967	0.9823	0.0248	0.5503	0.3029
Zoo	NB	0.9312	0.9274	0.9345	0.9538	0.1472	1.6172	0.6923
	SVM	0.9463	0.9523	0.9453	0.9618	0.0289	1.5669	0.6239
	KNN	0.9848	0.9851	0.9745	0.9819	0.0154	1.3587	0.6154

4.3 Performance analysis

The proposed work uses eight optimization techniques and three varied classifiers to analyze the optimal feature selection algorithm. The best techniques are selected for the proposed feature selection and classification process. The results obtained from the feature selection process are described in the following section. Table 10 represents the performance results of the proposed techniques using different datasets.

This Table shows the performance analysis of several feature selection techniques using metrics like mean, best, worst, standard deviation and computational time. The analysis is done by using seven different datasets. The result analysis shows that the SMA algorithm is better than other techniques, and the GWO algorithm is the second best among the other optimization techniques. The attained results of feature selection techniques in each dataset are illustrated in Fig. 2.

The obtained result analysis shows that the algorithms SMA and BGWO are superior to the other techniques. Thus, the proposed work introduces the hybrid optimized feature selection process using SMA and BGWO. This hybrid algorithm is compared with other optimization algorithms and is depicted in Table 11.

The hybrid techniques are compared with optimization algorithms like WOA, BAT, GWO and SMA. Using seven varied datasets, the analysis is performed. Each dataset contains several features, and the optimal features are selected using the mentioned algorithms. The comparison results prove that the hybrid approach gives improved results than the

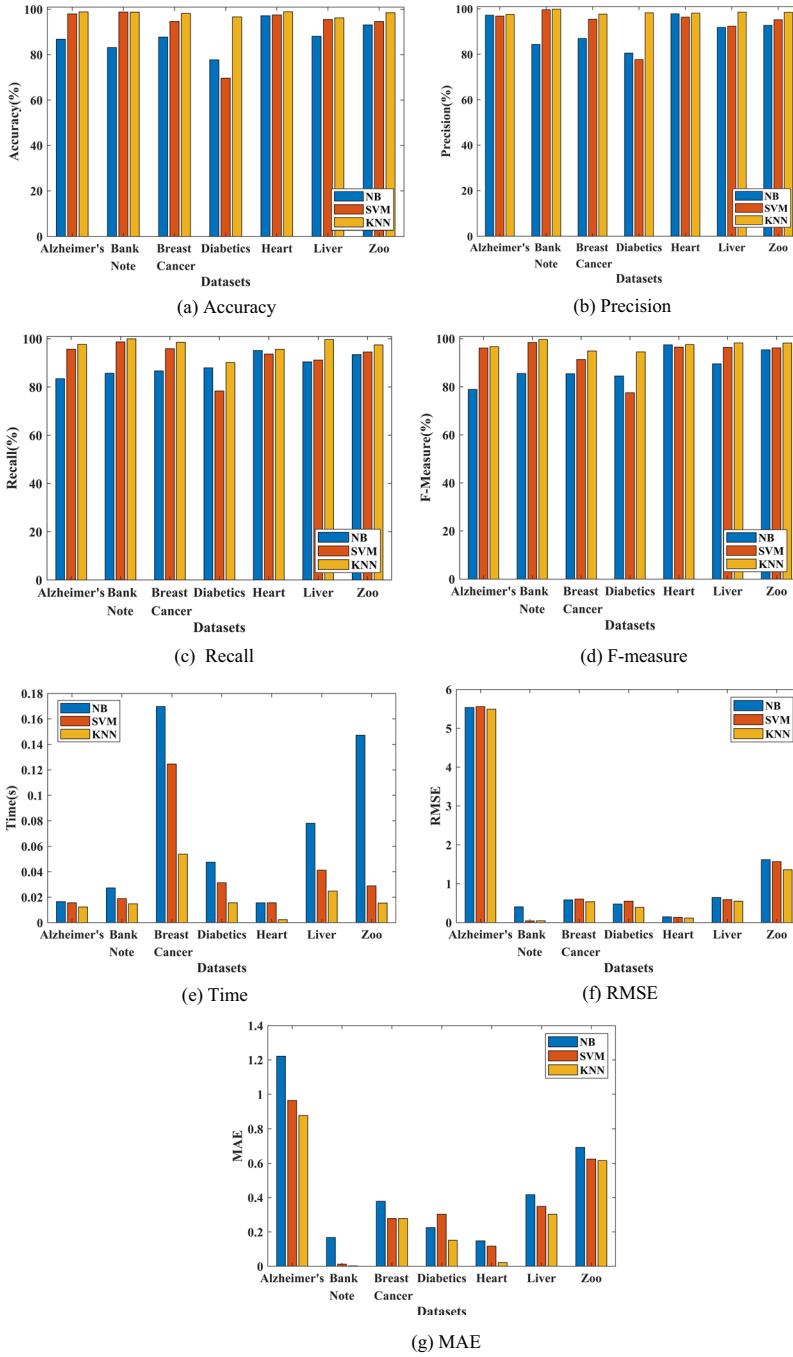


Fig. 5 Result analysis of different classifiers in terms of accuracy, precision, recall, F-measure, time, RMSE and MAE

Table 14 Comparing the performance of classifiers based on the SMA + BGWO approach using seven datasets

Data Set	CLASSIFIER	Accuracy	Precision	Recall	F-Measure	Time	RMSE	MAE
Alzheimer's	NB	0.8772	0.9732	0.8435	0.7989	0.0154	5.3912	1.2202
	KNN	0.9902	0.97653	0.9865	0.9765	0.0114	5.4332	0.7657
	SVM	0.97987	0.9693	0.9764	0.9699	0.0152	5.5234	0.9711
Banknote	NB	0.8726	0.8528	0.8689	0.8621	0.0221	0.3905	0.16715
	KNN	0.9899	0.9987	0.9998	0.9975	0.0139	0.0392	0.0018
	SVM	0.9897	0.9969	0.9888	0.9872	0.0181	0.0424	0.0121
Breast Cancer	NB	0.8832	0.8756	0.8754	0.8621	0.1476	0.5246	0.3272
	KNN	0.9887	0.9788	0.9896	0.9545	0.0499	0.5623	0.2671
	SVM	0.9561	0.9586	0.9635	0.9216	0.1231	0.6011	0.2723
Diabetes	NB	0.7823	0.8194	0.8852	0.8576	0.0415	0.4724	0.2153
	KNN	0.9783	0.9914	0.9233	0.9562	0.0121	0.3719	0.1493
	SVM	0.7014	0.7881	0.7899	0.7784	0.0304	0.5411	0.3010
Heart	NB	0.9794	0.9799	0.9598	0.9792	0.0124	0.1412	0.1331
	KNN	0.9917	0.9882	0.9626	0.9831	0.0021	0.1121	0.0217
	SVM	0.9788	0.9692	0.9435	0.9678	0.0152	0.1359	0.1167
Liver	NB	0.8893	0.9236	0.9144	0.9084	0.0776	0.6455	0.4081
	KNN	0.9793	0.9951	0.9987	0.9876	0.0227	0.5412	0.2927
	SVM	0.9643	0.9335	0.9186	0.9701	0.0407	0.5878	0.3301
Zoo	NB	0.9392	0.9384	0.9395	0.9587	0.1466	1.6165	0.6903
	KNN	0.9918	0.9901	0.9825	0.9889	0.0151	1.3456	0.6123
	SVM	0.9503	0.9578	0.9493	0.9625	0.0272	1.5588	0.6221

other techniques. Figure 3 represents the performance comparison of the proposed hybrid algorithm with other optimization techniques.

The comparison analysis states that the hybrid technique results better than the other optimization techniques. The hybrid algorithm selects the optimal set of features in reduced computational time. Also, these hybrid approaches are very easy to implement and can ignore local optima. Thus, this makes the algorithm attain enhanced results over the other techniques. To detect the efficacy of the proposed hybrid algorithm, the obtained results are compared with some of the previous existing hybrid approaches. Table 12 shows the performance comparison of the proposed hybrid optimization algorithm with other hybrid approaches.

The proposed hybrid approaches are compared with the existing hybrid optimization techniques like BAT-PSO [37], ALO-GWO [38], GWO-WOA [39] and SMA-FA [40]. These existing hybrid algorithms fail to provide better feature selection results because of the high computational time and slow convergence rate. Also, these approaches easily fall into local optima issues, which may affect the algorithm's performance. This drawback can be overcome in the proposed hybrid SMA + BGWO approach. This hybrid approach provides improved results regarding best, worst, standard deviation, mean and computational time. Figure 4 mentions the comparative analysis of the proposed hybrid algorithm with existing techniques in terms of several performance metrics.

This result analysis shows that the proposed hybrid approaches are more effective for feature selection than previous hybrid approaches. The proposed SMA + BGWO obtains more excellent results for each performance metric than the other hybrid algorithms.

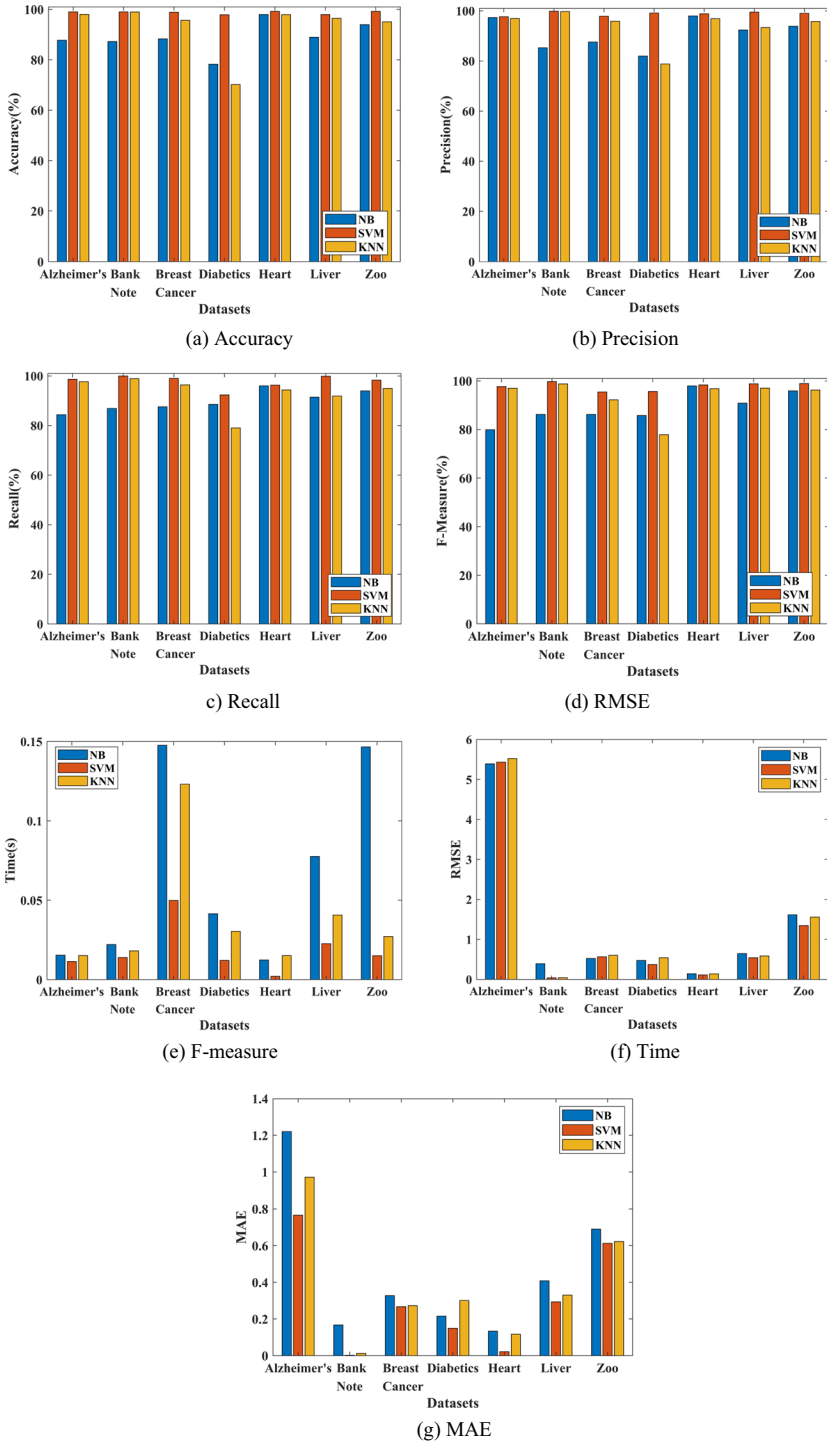


Fig. 6 Different classification performance based on the proposed hybrid SMA+BGWO approach using seven datasets

Table 15 Performance of KNN using a large dataset

Dataset	Accuracy	Precision	Recall	F-measure	MAE	RMSE
CICDDoS2019	0.9983	0.9983	0.9987	0.9985	0.0017	0.0412
CICMalDroid2020	0.9930	0.9969	0.9828	0.9987	0.0016	0.0476

Table 16 Performance of KNN using the medium dataset

Dataset	Accuracy	Precision	Recall	F-measure	MAE	RMSE
Heart	0.9925	0.9950	0.9950	0.9925	0.0075	0.0866
Wine quality	0.9929	0.9943	0.9943	0.9929	0.0071	0.0845
Car sales	0.9930	0.9940	0.9940	0.9930	0.0070	0.0837

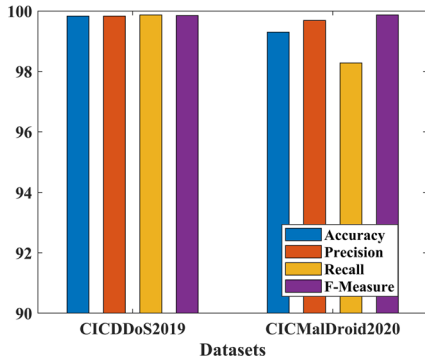
Table 17 Performance of KNN using the low dataset

Dataset	Accuracy	Precision	Recall	F-measure	MAE	RMSE
Bank note	0.9930	0.9970	0.9933	0.9849	0.0070	0.0703
Seeds	0.9959	0.9931	0.9947	0.9444	0.0059	0.0593
Prima Indians diabetes	0.9950	0.9960	0.9960	0.9950	0.0050	0.0701

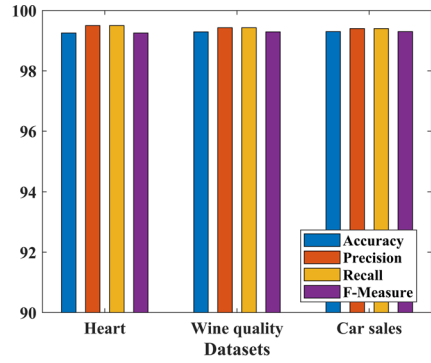
The selected features from these hybrid feature selection approaches are given as the input of the three classifiers: SVM, NB and KNN. Depending on this analysis, the best classifier uses varied performance metrics like accuracy, precision, recall, F-measure, RMSE, time and standard deviation. Table 13 compares the classification performance of different classifiers using seven varied datasets.

The analysis uses seven datasets with three varied classifiers: SVM, NB and KNN. The KNN approach provides better results for each metric than the SVM and NB methods. The SVM has an overlapping issue; hence, it is difficult to afford better classification results and takes a large amount of training time. Moreover, the SVM method cannot support large dimensionality data processing. Due to the high computational complexity, the existing NB cannot provide improved classification results. Thus, these approaches attain reduced results as compared with KNN. The KNN method is very easy to handle and provides a reduced error rate during classification. Also, the KNN approach can process large data dimensionality and improve the system's performance. Figure 5 shows the results based on classification using seven types of datasets.

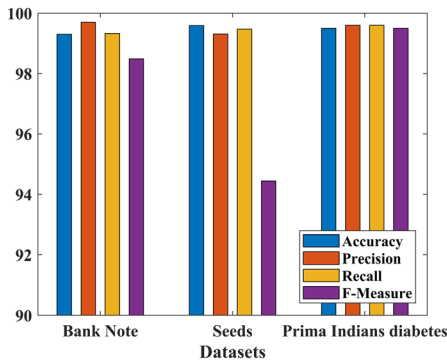
The performance analysis of three types is classifier analyzed using seven varied datasets. Compared with other approaches, the KNN method is effective for each dataset. For every metric, the KNN method attains improved classification results. According to this analysis, KNN is chosen as the best classifier. The proposed work integrates the s-shaped curve to the SMA+GWO algorithm to enhance the hybrid feature selection process and is termed SMA+BGWO. This process can be performed using both normal data and high data dimensionality. Table 14 represents the performance comparison of classification approaches.



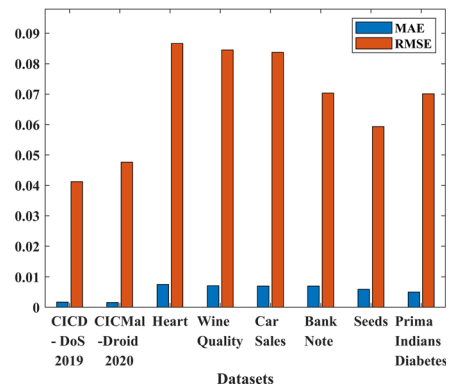
(a) Large dataset



(b) Low dataset

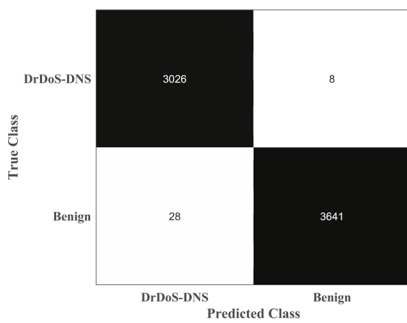


(c) Medium dataset

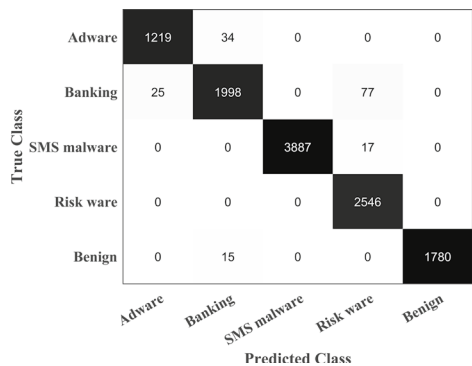


(d) RMSE and MAE

Fig. 7 Performance comparison of KNN classification with SMA+BGWO approach using high dimensionality datasets



(a) CICDDoS2019



(b) Banking note dataset

Fig. 8 Confusion matrix for CICDDoS2019 and Bank note dataset

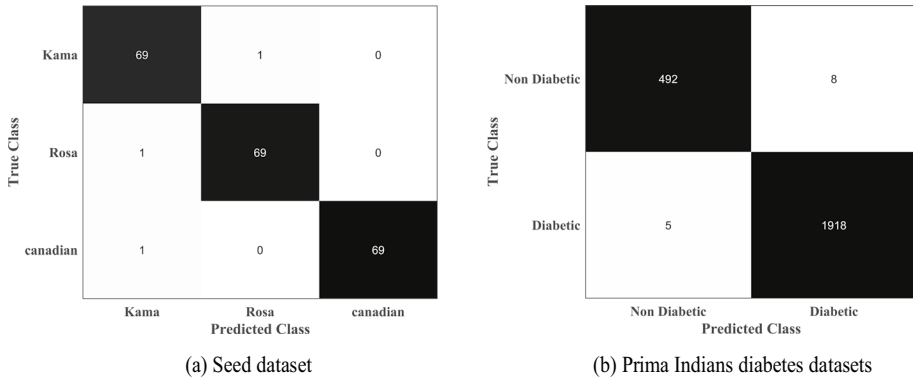


Fig. 9 Confusion matrix for Seed and Prima Indians diabetes dataset

Fig. 10 ROC analysis of the proposed hybrid approach

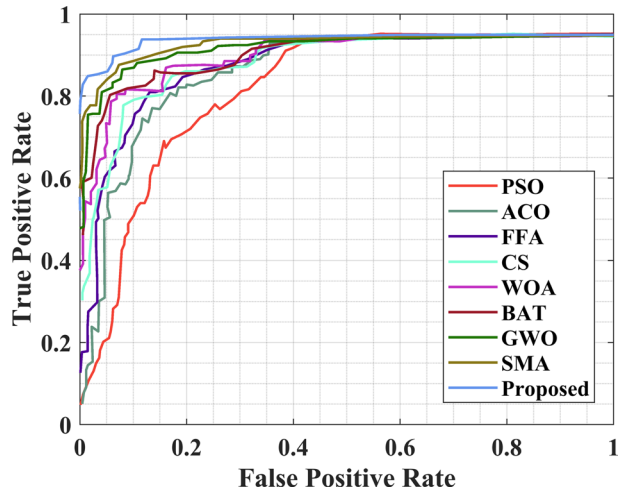
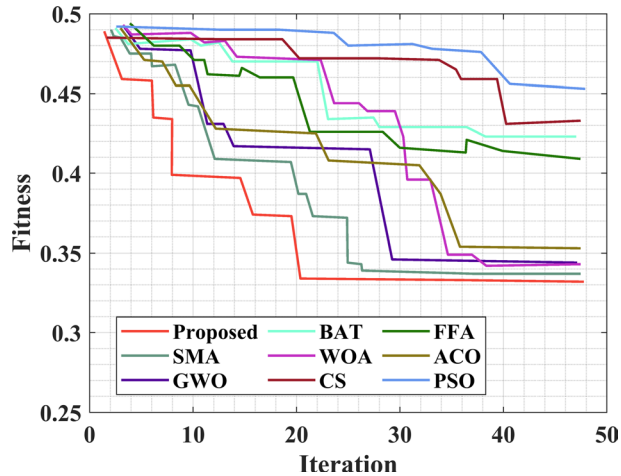


Fig. 11 Convergence analysis for the proposed hybrid optimization model



Using SMA + BGWA, the performance is enhanced in each performance metric. Different high dimensional datasets do the analysis, and the classification performance is improved for SVM, NB and KNN. Compared to two other classifiers, the KNN method attains improved outcomes. Figure 6 represents the performance comparison of three classifiers based on the SMA + BGWO approach.

The result analysis proves that the KNN method is more effective than the other two classification techniques. The proposed SMA + BGWO approach is also suitable for the large dimensionality of data. Tables 15, 16 and 17 represent the classification performance of KNN using large, medium and low datasets.

The performance analysis of KNN classification is performed by evaluating metrics like accuracy, precision, recall, F-measure, MAE and RMSE. For every dataset, the KNN classifier attains enhanced results in each metric. The hybrid SMA + BGWO model with KNN achieves effective results compared to the previous results. Figure 7 represents the performance comparison of KNN classifiers based on the SMA + BGWO approach dataset.

The KNN classifier achieves better results for each metric in different high dimensionality datasets. The results show that the hybrid optimization approach is highly suitable for the feature selection. Also, the proposed techniques are more applicable to feature selection and classification mechanisms. The confusion matrix for the datasets, namely CICDDoS2019 and bank notes, are depicted in Fig. 8.

The graphic above shows that the confusion matrix is evaluated for two datasets: CICDDoS2019 and the banknote dataset. The proposed model can predict the results for CICDDoS2019 more accurately than other existing models by considering two classes, DrDOS-DNS and Benign. The accuracy rate of finding these classes is 99.47%. Similarly, four classes, such as adware, banking, SMS malware, riskware and harmless, are classified from the banknote dataset with an accuracy of 99.05%. The confusion matrix for the Seeds and Prima Indians Diabetes datasets is shown in Fig. 9.

The above graphic shows that the confusion matrix is evaluated for two data sets: the diabetes data set “Seed” and “Prima Indians”. The proposed model can predict the results for the seed dataset more accurately than other existing models by considering the three classes: Kama, Rosa and Canadian. The accuracy rate of finding these classes is 99.57%. Similarly, from the Prima Indians diabetes dataset, two classes such as diabetics and non-diabetics, are classified with an accuracy of 99.52%. To better evaluate the proposed hybrid model, an ROC analysis is performed and shown in Fig. 10.

The ROC analysis of the proposed FS approach shows that the model leads to higher performance in both classification and prediction. In addition, the ROC lines are plotted between the true positive data rate and the false positive data rate. The existing models provide a linear constant true positive rate of 0.70 to 0.90. Even the proposed model provides a true positive rate of 0.83 to 0.95 for the classification process. The convergence plot for the proposed hybrid approach is shown in Fig. 11.

To measure the system performance and analyze the more cost-effective function, the proposed method is evaluated using convergence analysis compared to existing models. The models included in the comparison are PSO, ACO, FFA, CS, GWO, BAT, WOA and SMA. Using the 0–50 iterations of fitness values plotted for the proposed approach shows that it decreases the cost when it reaches the 20th iteration. Therefore, from the analysis, it is evident that the proposed SMA + BGWO model converges very quickly, showing better performance compared to other existing algorithms.

5 Conclusion

This paper executes several feature selection techniques based on the results, and the best algorithm is selected for further process. To solve the problems in the feature selection process, the proposed work combines SMA with BGWO algorithms. In the beginning, the data from the various high dimensional datasets are pre-processed, and data cleaning is performed. The pre-processed images are subjected to the hybrid optimization technique. In this, the optimal features are selected, which smoothens the classification process and increases the classification accuracy. For classification, the KNN approach is utilized. The performance of the KNN method is computed by measuring various performance metrics. The proposed model attains an accuracy of 0.9902 in the Alzheimer's dataset, the accuracy of the Banknote dataset is 0.9899, the accuracy of the Breast Cancer dataset is 0.9887, and the accuracy of the Diabetes dataset is 0.9783. The accuracy obtained in the Heart dataset is 0.9882, the Liver dataset is 0.9793, and the ZOO dataset is 0.9918. Using a large dimensional dataset, the proposed hybrid optimization with the KNN method attains the accuracy of the CICDDoS2019 dataset is 0.9983, achieved an accuracy of CICMalDroid2020 is 0.9930, a Heart dataset obtains an accuracy value of 0.9925, and the Wine quality dataset achieved 0.9929 of accuracy. The Car sales dataset attains 0.9930 accuracy, the banknote dataset attains the accuracy range of 0.9930, the seed dataset attains 0.9959 accuracy, and the Prima Indians dataset obtains 0.9950 accuracy value. The simulation results show that the developed model is more suitable for feature selection and classification. However, the proposed hybrid optimization is more computationally expensive than other recent population-based optimization algorithms. Thus, it should be solved in the future by adopting less computational optimization methods. Additionally, future work is suggested to manage the multi-objective formulation of the feature selection issue and introduce a multi-objective optimization technique to determine the Pareto best solution for feature selection issues. Also, the parameters in the optimization techniques will be adaptively updated to attain enhanced searching ability. Further, promoting the proposed optimization by utilizing evolutionary operators in which high scale optimization tasks will be solved.

Authors contributions All authors have equal contributions to this work.

Funding No funding is provided for the preparation of the manuscript.

Data availability Data sharing is not applicable to this article.

Declarations

Ethical approval This article does not contain any studies with human participants or animals performed by authors.

Consent to participate All authors have agreed to participate in this submitted article.

Consent to publish All the authors involved in this manuscript give full consent for publication of this submitted article.

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Venkatesh B, Anuradha J (2019) A review of feature selection and its methods. *Cybern Inf Technol* 19(1):3–26
2. Bolon-Canedo V, Remeseiro B (2020) Feature selection in image analysis: a survey. *Artif Intell Rev* 53(4):2905–2931
3. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: A new perspective. *Neuro-computing* 300:70–79
4. Nie F, Yang S, Zhang R, Li X (2018) A general framework for auto-weighted feature selection via global redundancy minimization. *IEEE Trans Image Process* 28(5):2428–2438
5. Raj RJ, Shobana SJ, Pustokhina IV, Pustokhin DA, Gupta D, Shankar KJ (2020) Optimal feature selection-based medical image classification using deep learning model in internet of medical things. *IEEE Access* 8:58006–58017
6. Kang X, Xiang X, Li S, Benediktsson JA (2017) PCA-based edge-preserving features for hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 55(12):7140–7151
7. Ghogh B, Samad MN, Mashhadi SA, Kapoor T, Ali W, Karray F, Crowley M (2019) Feature selection and feature extraction in pattern analysis: a literature review. *arXiv preprint arXiv:1905.02845*
8. Hussien AG, Hassanien AE, Houssein EH, Bhattacharyya S, Amin M (2019) S-shaped binary whale optimization algorithm for feature selection. *Recent trends in signal and image processing*. Springer, Singapore, pp 79–87
9. Tao Z, Huiling L, Wenwen W, Xia Y (2019) GA-SVM based feature selection and parameter optimization in hospitalization expense modeling. *Appl Soft Comput* 75:323–332
10. Zhang Y, Gong DW, Sun XY, Guo YN (2017) A PSO-based multi-objective multi-label feature selection method in classification. *Sci Rep* 7(1):1–12
11. Adeli A, Broumandnia A (2018) Image steganalysis using improved particle swarm optimization based feature selection. *Appl Intell* 48(6):1609–1622
12. Rostami M, Forouzandeh S, Berahmand K, Soltani M (2020) Integration of multi-objective PSO based feature selection and node centrality for medical datasets. *Genomics* 112(6):4370–4384
13. Mafarja M, Aljarah I, Heidari AA, Faris H, Fournier-Viger P, Li X, Mirjalili S (2018) Binary dragonfly optimization for feature selection using time-varying transfer functions. *Knowl-Based Syst* 161:185–204
14. Parvathavarthini S, Deepa D (2021) A hybrid artificial neural network classifier based on feature selection using binary dragonfly optimization for breast cancer detection. In *IOP Conf Ser: Mater Sci Eng* 1055(1):012107 (IOP Publishing)
15. Chantar H, Tubishat M, Essgaer M, Mirjalili S (2021) Hybrid binary dragonfly algorithm with simulated annealing for feature selection. *SN Comput Sci* 2(4):1–11
16. Mafarja M, Mirjalili S (2018) Whale optimization approaches for wrapper feature selection. *Appl Soft Comput* 62:441–453
17. Agrawal RK, Kaur B, Sharma S (2020) Quantum based whale optimization algorithm for wrapper feature selection. *Appl Soft Comput* 89:106092
18. Kumar A, Gandhi CP, Liu X, Liu Y, Zhou Y, Kumar R, Xiang J (2022) A novel health indicator developed using filter-based feature selection algorithm for the identification of rotor defects. *Proc Inst Mech Eng Part O: J Risk Reliab* 236(4):529–541
19. Allam M, Malaiyappan N (2020) Wrapper based feature selection using integrative teaching learning based optimization algorithm. *Int Arab J Inf Technol* 17(6):885–894
20. Madasu A, Elango S (2020) Efficient feature selection techniques for sentiment analysis. *Multimed Tools Appl* 79:6313–6335
21. Deng X, Li Y, Weng J, Zhang J (2019) Feature selection for text classification: A review. *Multimed Tools Appl* 78:3797–3816
22. Sathiyabhama B, Kumar SU, Jayanthi J, Sathiya T, Ilavarasi AK, Yuvarajan V, Gopikrishna K (2021) A novel feature selection framework based on grey wolf optimizer for mammogram image analysis. *Neural Comput Appl* 33(21):14583–14602
23. Shanthi S, Rajkumar N (2021) Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods. *Neural Process Lett* 53(4):2617–2630
24. Asgarnezhad R, Monadjemi SA, Soltanaghaei M (2021) An application of MOGW optimization for feature selection in text classification. *J Supercomput* 77(6):5806–5839
25. Ghosh M, Guha R, Sarkar R, Abraham A (2020) A wrapper-filter feature selection technique based on ant colony optimization. *Neural Comput Appl* 32(12):7839–7857
26. El-Kenawy ES, Eid M (2020) Hybrid gray wolf and particle swarm optimization for feature selection. *Int J Innov Comput Inf Control* 16(3):831–844

27. Pathak Y, Arya KV, Tiwari S (2019) Feature selection for image steganalysis using levy flight-based grey wolf optimization. *Multimed Tools Appl* 78:1473–1494
28. Singh LK, Khanna M, Thawkar S, Singh R (2022) Collaboration of features optimization techniques for the effective diagnosis of glaucoma in retinal fundus images. *Adv Eng Softw* 173:103283
29. Khanna M, Singh LK and Garg H (2023) A novel approach for human diseases prediction using nature inspired computing & machine learning approach. *Multimed Tools Appl* 1–37
30. Singh LK, Khanna M (2023) A novel enhanced hybrid clinical decision support system for accurate breast cancer prediction. *Measurement* 221:113525
31. Singh LK, Khanna M, Garg H and Singh R (2023) Emperor penguin optimization algorithm-and bacterial foraging optimization algorithm-based novel feature selection approach for glaucoma classification from fundus images. *Soft Comput* 1–37
32. Khan AAR, Nisha SS, Sathik MM (2022) Hybrid slime mould-Grey wolf optimization algorithm for efficient feature selection. *Int J Health Sci I*:7657–7663
33. Ge D, Zhang Z, Kong X, Wan Z (2022) Extreme learning machine using bat optimization algorithm for estimating state of health of lithium-ion batteries. *Appl Sci* 12(3):1398
34. Chitara D, Niazi KR, Swarnkar A, Gupta N (2018) Cuckoo search optimization algorithm for designing of a multimachine power system stabilizer. *IEEE Trans Ind Appl* 54(4):3056–3065
35. Aydilek IB (2018) A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems. *Appl Soft Comput* 66:232–249
36. Li S, Chen H, Wang M, Heidari AA, Mirjalili S (2020) Slime mould algorithm: A new method for stochastic optimization. *Futur Gener Comput Syst* 111:300–323
37. Tawhid MA, Dsouza KB (2018) Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. *Appl Comput Inf* 16(1/2):117–136
38. Zawbaa HM, Emary E, Grosan C, Snasel V (2018) Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach. *Swarm Evol Comput* 42:29–42
39. Mafarja M, Qasem A, Heidari AA, Aljarah I, Faris H, Mirjalili S (2020) Efficient hybrid nature-inspired binary optimizers for feature selection. *Cogn Comput* 12(1):150–175
40. Ewees AA, Abualigah L, Yousri D, Algamal ZY, Al-Qaness MA, Ibrahim RA, Abd Elaziz M (2021) Improved slime mould algorithm based on firefly algorithm for feature selection: a case study on QSAR model. *Eng Comput* 1–15

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

A. Ameer Rashed Khan¹ · S. Shajun Nisha²

✉ A. Ameer Rashed Khan
ameerkhan.a1694@gmail.com

¹ Department of Data Science, The New College, Royapettah, Chennai, Tamil Nadu 600014, India

² PG & Research Department of Computer Science, Sadakathullah Appa College, Tirunelveli, Tamil Nadu 627011, India