DOI: 10.24996/ijs.2026.67.4.40

ISSN: 0067-2904

A Hybrid Rough Set-Based Binary Grasshopper Optimization Algorithm for Feature Selection

Salam Khalaf Abdullah[1], Wisam Ali Mahmood[2], Layth Kamil Adday Almajmaie[3], Jumana Waleed[4], Maisa՚ a Abid Ali Khodher[3*]

[1]*Department of Computer Engineering, Al. Nukhba University, Baghdad, Iraq*
[2]*Department of Computer Science, University of Technology-Iraq, Baghdad, Iraq*
[*3]*Department of Computer Engineering, University of Technology-Iraq, Baghdad, Iraq*
[4]*Department of Computer Science, College of Science, University of Diyala, Diyala, Iraq*

Received: 3/3/2024 Accepted: 30/4/2025 Published: 30/4/2026

Abstract

Feature Selection (FS) is a technique that removes redundant and unnecessary characteristics from the original data to identify the smallest subset of features. Its goal is to make classification algorithms more efficient. Rough set theory (RST) offers a reliable route to feature selection; however, it resorts to comprehensive searches to find all subsets of features and dependence to assess them. However, due to its high cost, the entire search may not be viable for huge data sets. As a result, meta-heuristic algorithms, particularly Nature-Inspired Algorithms, are commonly employed to substitute the RST reduction step. The Hybrid Rough Set based Binary Grasshopper Optimization Algorithm (HRBGOA) was proposed as a FS approach for given datasets using BGOA with Rough Set to achieve significant Size Reduction Proportion (SR%) without significantly lowering classification accuracy compared to the total number of features in a smaller number of iterations. Compared to the Binary Grasshopper Optimization Algorithm (BGOA) and Particle Swarm Optimization (PSO) techniques, the experimental findings reveal that HRBGOA produced improved FS in seven datasets.

Keywords: Feature Selection, Nature-Inspired Algorithms, Rough Sets, Binary Grasshopper Optimization Algorithm, Classification.

خوارزمية تحسين الجندب الثنائية الهجينة القائمة على المجموعة لاختيار الميزة

سلام خلف عبدالله[1], وسام علي محمود[2], ليث كامل عداي المجمعي[3], جمانة وليد[4], ميساء عبد علي خضر[3*]

[1]قسم هندسة الحاسوب , كلية النخبة , بغداد, العراق
[2]قسم علوم الحاسوب , الجامعة التكنلوجية , بغداد, العراق
[3]قسم هندسة الحاسوب ,الجامعة التكنلوجية , بغداد, العراق
[4]قسم علوم الحاسوب، كلية العلوم، جامعة ديالى، ديالى، العراق

الخلاصة

تحديد الميزة (FS) هو أسلوب يزيل الخصائص الزائدة وغير الضرورية من البيانات الأصلية لتحديد أصغر مجموعة فرعية من الميزات. هدفها هو جعل خوارزميات التصنيف أكثر كفاءة. توفر نظرية

*Email: Maisaa.a.khodher@uotechnology.edu.iq

المجموعات التقريبية (RST) طريقًا موثوقًا لاختيار الميزات، إلا أن الأمر يتطلب اللجوء إلى عمليات بحث شاملة للعثور على جميع المجموعات الفرعية من الميزات والاعتماد عليها لتقييمها .ومع ذلك، نظرًا لتكلفته العالية، قد لا يكون البحث بأكمله قابلاً للتطبيق بالنسبة لمجموعات البيانات الضخمة .ونتيجة لذلك، يتم استخدام الخوارزميات الفوقية الإرشادية، وخاصة الخوارزميات المستوحاة من الطبيعة، بشكل شائع لتحل محل خطوة تخفيض RST. تم اقتراح خوارزمية تحسين الجندب الثنائية القائمة على مجموعة Hybrid Rough (HRBGOA)كنهج لاختيار الميزات لمجموعات بيانات معينة باستخدام BGOA مع Rough Set لتحقيق نسبة كبيرة لتخفيض الحجم (SR%) دون تقليل دقة التصنيف بشكل كبير مقارنة بالعدد الإجمالي الميزات في عدد أقل من التكرارات .عند مقارنتها بخوارزمية تحسين الجندب الثنائي (BGOA) وتقنيات تحسين سرب الجسيمات(PSO) ، تكشف النتائج التجريبية أن HRBGOA أنتجت FS محسّنًا في سبع مجموعات بيانات.

1. Introduction

As data acquisition technology becomes more widely available and used, data collection is increasing exponentially. Structured datasets are becoming increasingly common, and their size and complexity are growing. This increases the need and demand for data mining algorithms to dynamically retrieve increasingly precise views from massive and complex databases [1]. Classification is an essential tool for extracting useful information. It frequently employs data mining techniques to categorize every record in a hierarchical dataset.

Pattern recognition and classification systems, such as automatic numeral or iris recognition systems, are facilitated by feature choice [2] [3]. These procedures' main objective is to select a limited subset of features from a problem domain that accurately reflects the natural composition. It improves the predicted precision of algorithms by reducing the number of features and removing unnecessary, noisy, and repetitive variables. Feature selection technology has been widely used to address classification problems in fields as diverse as bioinformatics, image processing, data mining, pattern recognition, and medical diagnostics. The dimensionality constraint, caused by the large number of features in the datasets, is a fundamental difficulty for classification systems. This problem improves the classification model's computing time, overcomes the overfitting problem, and reduces classification precision.

Overfitting is thought to be caused primarily by duplicated characteristics. Substantially, Feature selection (FS) addresses dimensionality by selecting the most relevant features from a dataset to begin classification [4]. To address this issue, FS has been implemented. The fundamental goal of FS is to find the smallest subset of features in a problem domain while still accurately reflecting the natural elements. Many approaches were used to achieve feature selection. This includes filter, embedded, and wrapper approaches [5]. The filter-based FS approaches are computationally effective and independent of classifiers. Hence, they do not interact with classifiers. These approaches are categorized into parametric filters (such as Chi-squared and ANOVA) and non-parametric filters (such as Information gain and ReliefF). Embedded-based FS approaches are classifier-based, efficient for medium-sized datasets, and select features during the training process. However, their outcomes might not generalize to other classifiers.

Tree-based approaches (such as Extreme Gradient Boosting and Random Forest), Elastic Net, and Lasso Regression are well-known examples of embedded approaches. Wrapper-based FS approaches are also classifier-based, in which the selected subsets of features are assessed via training classifiers to evaluate their performance. Wrapper approaches utilize various search strategies (such as backward elimination, forward selection, and meta-heuristic algorithms) for exploring the feature space. Although these approaches are considered feature interactions, they are computationally expensive since they entail training the classifiers repeatedly throughout the FS process [6]. Generally, implementing feature selection approaches with machine learning methods could be computationally expensive and necessitate further expertise and resources [7].

Diverse meta-heuristic algorithms have materialized by inspiring numerous disciplines like sociology, physics, and biology to explore optimal prospective solutions [8] [9]. Until now, many FS approaches have looked for the ideal subset using Meta-heuristic algorithms. Exploration and exploitation are the two guiding principles of these algorithms. By maintaining and fusing these two ideas, the algorithm can be strengthened. Regretfully, these two ideas do not properly interact with other algorithms. For instance, the genetic algorithm incorporates these two ideas but is unable to create a meaningful relationship between them [10].

The grasshopper optimization method has gained greater attention due to its features, which include rapid synchronization and straightforward implementation. The main idea behind rough set-based feature selection is to generate all feasible feature reductions before choosing the lowest cardinality. Attempts have been made to integrate the Rough Set (RS) method with bio-themed algorithms to improve results [11]. All of these methods have major flaws: First, they have poor convergence when applied to high-dimensional datasets because bio-inspired approaches' preliminary and global processes do not span the entire search space. Second, because their RST-based goal function is only effective for nominal datasets, it is inefficient for numerical and mixed datasets.

This paper proposes a more effective filter FS by inventing Hybrid Binary Grasshopper Optimization (HBGOA), a new objective function based on frequent values and Rough Set Theory (RST), to solve the aforementioned constraints. To improve HBGOA's convergence effectiveness, this work established a new starting mechanism that spans much of the search area and an emerging updating technique with more effective convergence for both low- and high-dimensional datasets. This increases the effectiveness of the proposed mechanism for all datasets, including nominal, mixed, and numerical data. This strategy aims to reduce the number of features chosen while improving or maintaining classification precision in all of these high-dimensional datasets.

2. Related Work

Because swarm-based algorithms are simple and effective in global optimization, they are frequently suggested in the literature as solutions to difficult and complex optimization problems in various areas of search. Over the past ten years, there has been a sharp rise in the quantity of these algorithms [12] [13].

PSO is just one of many swarm-based optimization techniques that have been proposed in the literature [14], Ant Colony Optimization (ACO) [15], Grey Wolf Optimization (GWO) [16], Ant Lion Optimization (ALO) [17], Moth-Flame Optimization (MFO) [18], and Whale Optimization Algorithm (WOA) [19]. Nearly all of the methods mentioned above were first suggested for situations involving continuous optimization. After that, they were binarized to be used for binary optimization problems such as feature selection in data classification, which they perform better than should be expected [13]. Binary Particle Swarm Optimization (PSO) and its variants have been extensively employed in the FS problem. Moradi and Gholampour [20] suggested a hybrid PSO with a local search strategy to identify the feature subset with a lower correlation. Zhang et al. [21] used filter-based bare-bone PSO. In this work, two filter-based strategies are proposed (average of the mutual data and feature redundancy) to improve the possibility of exploitation of the swarm. A Competitive Swarm Optimizer (CSO) for solving high-dimensional FS problems is suggested by Gu et al. [22]. Xue et al. [23] presented initialization and upgrading procedures for the original PSO. A hybrid binary bat-enhanced particle swarm optimization algorithm called HBBEPSO was proposed by Tawhid and Dsouza [24]. The authors combined the bat algorithm and the PSO to solve the FS problem, exploiting the best aspects of both these algorithms. Wah et al. [25] demonstrated an FS and classification system in which the wrapper approaches chose more substantial features in contrast to the filter approaches; hence, the information gain-based FS did not perform well with continuous features. This system utilized three datasets from the UCI Machine-Learning Repository, confirming that the sequential backward elimination was the superior wrapper approach.

3. Rough Set Theory

Rough set theory (RST) was suggested by Pawlak [26] to tackle data imprecision, ambiguity, and inconsistency. Scholars have employed RST in several domains for various reasons. First, it helps find hidden data patterns. Second, it reduces data without more data. Third, it's straightforward. Fourth, it allows data-only importance evaluation [27]. The RST's approximation space is an ordered pair $\mathbf{I}=(\mathbf{U},\mathbf{R})$, where $\mathbf{I}$: Information System; $\mathbf{U}$: Nonempty Set of Objects.; $\mathbf{R}$: Equivalence Relation on $\mathbf{U}$, called the Indiscernibility Relation. For each $\mathbf{x}\in\mathbf{U}$, let $[\mathbf{X}]_{\mathbf{R}}$ denote the equivalence class of $\mathbf{R}$, which includes $\mathbf{x}$. For each $\mathbf{X}$, $\mathbf{X}\subseteq\mathbf{U}$, $\mathbf{X}$ is represented in $\mathbf{I}$ by a pair of sets- its lower and upper approximation in $\mathbf{I}$, defined respectively as $RX = \{X \in U[X]R \complement X$.The terms reduce and core are interchangeable in RST [28]. A collection of characteristics is reduced to preserve partition. A reduct is the smallest set of characteristics that allows the same classification of global components while considering entire attribute sets. In an information system $\mathbf{I}=(\mathbf{U},\mathbf{A})$, let $\mathbf{B}\subseteq\mathbf{A}$ and $\mathbf{a}\in\mathbf{B}$ be defined as the $\mathbf{U}$ universe of objects, the $\mathbf{A}$ set of characteristics, and the $\mathbf{R(B)}$ binary relation respectively.

- A can be omitted in $\mathbf{B}$ if $\mathbf{R(B)}=\mathbf{R(B-\{a\})}$ conditions are met; otherwise, $\mathbf{a}$ is indispensable in $\mathbf{B}$.
- A set $\mathbf{B}$ is considered independent if each and every one of its properties is required.

$\mathbf{B'}\subseteq\mathbf{B}$ is a reduct of $\mathbf{B}$ if $\mathbf{B'}$ it is $\mathbf{R(B')}=\mathbf{R(B)}$ and independent. All the attributes except the reduct are superfluous. The categorization cannot be damaged by removing duplicate characteristics. A dataset usually contains numerous reducts. The core is the confluence among reducts since it contains all essential attributes. If we define the core of $\mathbf{B}$ an information system $\mathbf{I}=(\mathbf{U},\mathbf{A})$ where $\mathbf{B}\subseteq\mathbf{A}$ as the set of all reducts of that system, then let $\mathbf{Red(B)}$ is the set $\mathbf{B}$ of all reducts, $\mathbf{core(B)}=\bigcap\mathbf{Red(B)}$.

Because each reduct is based on the core, every component of the core is a reduct. Consequently, because no part of the core can be removed without affecting the classification, it is the most important set of features. It is possible to define the positive and negative regions as follows, if we take the assumption that $\mathbf{P},\mathbf{Q}\subseteq\mathbf{R}$ there are equivalence relations over $\mathbf{U}$.

$$\mathbf{POS_P(Q)} = \bigcup_{\mathbf{X}\in\mathbf{U/Q}} \underline{\mathbf{P}}\mathbf{X} \tag{1}$$

$$\mathbf{NEG_P(Q)} = \mathbf{U} - \bigcup_{\mathbf{X}\in\mathbf{U/Q}} \overline{\mathbf{P}}\mathbf{X} \tag{2}$$

$$\mathbf{BND_P(Q)} = \bigcup_{\mathbf{X}\in\mathbf{U/Q}} \overline{\mathbf{P}}\mathbf{X} - \bigcup_{\mathbf{X}\in\mathbf{U/Q}} \underline{\mathbf{P}}\mathbf{X} \tag{3}$$

The set of all objects that have a 100% chance of being able to be sorted into one of the blocks of the partition is the positive area of the partition U/Q when viewed in relation to $P(POS_P(Q))$. Finding connections between characteristics is an essential aspect of attribute reduction. U/Q by means of P. For $P, Q \subseteq R$, Q depending on P what we mean by $k(0 \le K \le 1)$ denoted $P \Rightarrow_k Q$ we say that:

$$k = \gamma_p(Q) = \frac{|POS_P(Q)|}{|U|} \tag{4}$$

If $0 \le k \le 1$ is the event that is completely or partially Q dependent on P, and in the event that if $k=0$, Q not dependent on, then P [11].

4. Grasshopper Optimization Algorithm

The grasshopper is a type of insect. Grasshoppers are considered pests because they impede crop production and farming. Grasshoppers have the unusual ability to swarm in both their juvenile and adult stages. They can only leap and roll around like a cylinder in early infancy, eating any vegetables they come across. As adults, they develop wings and form a swarm in the air. The immature stage is distinguished by tepid swarm motion and short grasshopper steps. A wide range of motion and precise movement are required for maturity. The key characteristics of Grasshopper drive it to extract and investigate. As is well known, all natural-inspired approaches divide the search process into investigation and extraction. Throughout the investigation, agents are pushed to act quickly and frequently go local during extraction. In terms of mathematical formulation, Grasshopper achieves target seeking naturally as follows:

$$X_j = S_j + G_j + A_j \tag{5}$$

This equation depicts the j^{th} grasshopper's placement X_j. The equation between grasshoppers S_j act as a gravitational pull G_j that determines the wind's direction A_j depicting a grasshopper's position. We may apply the equation to generate random behavior, where r can arbitrarily fluctuate between 0 and 1:

$$X_j = r_1 S_j + r_2 G_j + r_3 A_j \tag{6}$$

S_j is the social interaction that is determined according to d_{ij} donates the distance between i^{th} grasshopper and j^{th} grasshopper.

$$S_i = \sum_{\substack{j=1 \\ j \ne i}}^{N} S(d_{ij})\, d_{ij} \tag{7}$$

d_{ij} represents the distance between i^{th} and j^{th} two grasshoppers and is calculated as $d_{ij} = |X_j - X_i|_i$ according to the S function, the distance between grasshoppers is mapped.

$$S(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{8}$$

f is a measure of gravity's intensity, which follows the general formula:

$$X_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} S(|x_j - x_i|)\frac{x_j - x_i}{d_{ij}} - ge_g + ue_W \tag{9}$$

Grasshoppers because they move on the ground, their position must not exceed a certain threshold, causing the equation to be adjusted:

$$X_i^d = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\frac{ub_d - lb_d}{2} s(|x_j^d - x_i^d|)\frac{x_j - x_i}{d_{ij}}\right) + T_d \tag{10}$$

The equation for a grasshopper's position is predicated upon the grasshopper's present position, goal location, and the position of the rest of them. C is a reduction factor that affects the secure, repelling, and gravity areas, an essential factor of the grasshopper optimization method. The latest value for this parameter can be found by using the equation.

$$c = c_{max} - l\frac{c_{max} - c_{min}}{L} \tag{11}$$

4.1 Binary Grasshopper Optimization Algorithm

Continuous solution values are used in traditional grasshopper optimization methods. An approach to feature selection that works like this will handle the search space as if it were a Boolean network. The difficulty of selecting or not selecting a feature can be solved with the help of the binary vector. In this method, the values zero and one represent the feature being picked and the feature not being selected, respectively. An algorithm known as the Grasshopper Optimization Algorithm (GOA) is translated into a binary version of the algorithm known as the Binary Grasshopper Optimization Algorithm (BGOA) as follows:

- Update the Parameter c,
- Nonlinear Cartography of v,
- Grasshoppers' New Location Updated.

In these Equations, rather than utilizing $G + A$, a T label is used. T is the best grasshopper that has ever been. The equation below shows the social interaction of grasshoppers' behavior in which: $X_i^d = cs_i + T_d$, where $S_1 = \sum_{i\neq j}^{N} \frac{c}{2} \cdot |x_j - x_i| d_j$.

The binary grasshopper optimization technique uses a nonlinear function V to move the social behavior of grasshoppers to a different function or space because the decision variable can only be between 0 and 1. The following equation is used to generate this function:

$$V_i = \left|\frac{2}{\pi}\tan^{-1}\left(\frac{10\pi}{2} cS_i\right)\right| \tag{12}$$

Last, we need to change the grasshoppers' location.

$$X_i = \begin{cases} T_i & p > v_i \\ \bar{X}_i & p < v_i \end{cases} \tag{13}$$

The suggested mathematical framework can conduct search space discovery and extraction. Nevertheless, there ought to be a system in place to modulate the amount of search factor investigation and extraction. The connection between the two notions of investigation and extraction is aided by modifying the parameter C, $c = cinf + ((1 - it^w)^{(1/w)}) * (1 - cinf)$.

The equation that follows represents a value that is extremely close to zero, which indicates the final value. This even considers the iterations of the algorithm . This algorithm can switch between inquiry and extraction by shifting the value of the parameter. Extraction attains better import in case $\mathbf{w}$ is quantitatively higher 0.5, while extraction gains the upper hand when $\mathbf{w}$ is smaller than 0.5.

5. Proposed Algorithm

To broaden the application of the GOA to binary issues, we present the Hybrid Rough Set based Binary Grasshopper Optimization Technique (HRBGOA) as an optimization algorithm to resolve feature selection issues raised as part of our work. When searching for the best feature subset for an application, FS-BCS applies BGOA as a search technique and uses an objective function based on rough sets to evaluate these candidate feature subsets. This suggested goal function picks the smallest number of chosen characteristics with the highest classification precision. The proposed HRBGOA is presented below

$$\mathbf{Obj\_Function(R)} = \gamma_{\mathbf{R}}(\mathbf{D}) * \frac{|\mathbf{C}| - |\mathbf{R}|}{|\mathbf{C}|} \tag{14}$$

There are |R| conditional features, which have been selected, and the number of selected features. D denotes the class label, and $\gamma_R(D)$ relies on the number of conditional features.

Algorithm: Hybrid Rough Set based Binary Grasshopper Optimization Algorithm (HRBGOA)
Input: Set of grasshoppers (N), Maximum and Minimum number of iterations for optimization. Output: Optimal grasshopper's binary position.

1. Create a population of grasshopper positions at random. $\mathbf{X_i\,(i=1,2,...,n)\ at\ from\ \{0,1\}}$.
2. Determine each grasshopper's fitness level. Find the finest solution T relied on the function of your fitness.
3. Let $\mathbf{ft=0}$;
4. $\mathbf{while(ft<Max_{ft})}$

$$\mathbf{Update\ c=C_{Max}} - 1\frac{\mathbf{C_{Max}} - \mathbf{C_{Min}}}{\mathbf{L}}$$

5. for (each grasshopper in the population) do

Normalize the distance between grasshoppers

Calculate Dist value and F(Dist)

$$\mathbf{Dist=c}\left(\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{c}{2}\mathbf{S}\left(\mathbf{G}\left(\left|\mathbf{X}_j^d - \mathbf{X}_i^d\right|\right)\right)\frac{\mathbf{x}_j - \mathbf{x}_i}{\mathbf{d}_{ij}}\right)$$

$$\mathbf{F}\left(\mathbf{Dist^d}\right)=\begin{cases}\mathbf{1,\ if\ Dist^d} > \dfrac{\mathbf{Max(Dist^d)\text{-}Min(Dist^d)}}{\mathbf{2}}\\ \mathbf{0,\ if\ Dist^d} < \dfrac{\mathbf{Max(Dist^d)\text{-}Min(Dist^d)}}{\mathbf{2}}\\ \mathbf{Else}\begin{cases}\mathbf{0,\ if\ \ Rand() < 0.5}\\ \mathbf{Else\ 1}\end{cases}\end{cases}$$

Update the position of the current grasshopper

$$\mathbf{X_{id}} = \begin{cases}\mathbf{T_d,\ if\ F(Dist^d) = 0}\\ \mathbf{Else\ \ |1\text{-}T_d|}\end{cases}$$

end for

6. Update the target T by the best position

ft=ft+1

end while Return T

6. Experimental Methodology

6.1 Datasets

The UCI Machine Learning Repository was used to get seven datasets to conduct a comprehensive comparison of the proposed approach with alternative techniques.

Table 1: Description of the dataset

Dataset	Features	Objects	Classes
Hepatitis	19	155	2
Mushroom	22	8124	2
Dermatology	33	366	6
Breast Cancer	9	699	2
Segment	19	1500	7
Glass Identification	10	214	6
HIGGS dataset	28	11,000,000	2

6.2 Performance Evaluation Measures

When a model can correctly predict new data based on the training datasets, it is referred to as having predictive accuracy. In data mining and machine learning, cross-validation is a common method for assessing model predictive performance. Data sets are typically split into two parts: the training set is used to develop a prediction model, and the test set is used to evaluate how well the model predicts. Cases below and above were cross validated for datasets with values of and, respectively, for datasets. We used a random number generator to divide each dataset into two sections: training and testing. Containing roughly all of the dataset's items, there is a training set and a test set with nearly all of the dataset's items. Accuracy, recall, precision, computational time, and reduction rate are among the assessment metrics employed.

- **Accuracy:** The most important factor to consider when determining how accurate a categorization algorithm is. It shows what percentage of the total set of test records has been correctly assigned to the proper categories. The following equation is used to determine the precision benchmark [29]:

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{15}$$

- **Precision:** It refers to the ratio between True Positives and all Positives [30]:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{16}$$

- **Recall:** It measures the capacity of the model to accurately locate True Positives [31]:

$$\text{Recall}=\frac{TP}{TP+FN} \tag{17}$$

- **Computational time:** It stands for the time needed to complete the application. It shows better algorithm performance in the case of lesser computation time.
- **Reduction Rate (SR):** It represents the reduction percentage in features, which is denoted as:

$$\text{Reduction rate}=1-\frac{ns}{nf} \tag{18}$$

This equation displays the algorithm's selection of features and the total number of attributes. As part of the experimentation, the classification performance of J48 Decision Trees and Support Vector Machines (SVM) is evaluated.

6.3 Results and Discussion

To assess classification performance, these studies look at classification accuracy (Acc) and the proportion of features removed compared to the total amount of features. The optimal method produces the highest SR percent while significantly lowering classification accuracy compared to the total number of features in the fewest number of repetitions. This is the best possible outcome. When the difference in precision is larger than five percent, it is seen as significant; when it is just under one percent, it is regarded as the same. Compared to the PSO and BGOA techniques, Table 2 indicates that HRBGOA produces a higher SR percent despite making no significant compromise in classification precision in fewer iterations. Figures 1 and 2 indicate that HRBGOA produces a higher accuracy than other FS approaches.

Table 2: Accuracy comparison of different approaches

Dataset	Approach	SR%	J48 Decision Tree Acc	SVM Acc
Hepatitis	PSO	78.94	84.7	85.4
	BGOA	78.94	86.9	87.3
	HRBGOA	79.13	89.6	90.2
Mushroom	PSO	72.73	98.6	98.7
	BGOA	72.73	98.7	98.9
	HRBGOA	72.81	99.1	99.7
Dermatology	PSO	61. 67	77.1	78.2
	BGOA	64.71	78.3	86.3
	HRBGOA	70.59	94.5	95.4
Breast Cancer	PSO	66.67	97.6	97.7
	BGOA	66.67	98.2	98.3
	HRBGOA	66.78	98.7	98.8
Segment	PSO	68.75	82.3	82.3
	BGOA	68.42	83.1	84.9
	HRBGOA	84.21	92.3	93.4
Glass Identification	PSO	64.31	80.7	85.3
	BGOA	64.84	83.6	86.1
	HRBGOA	65.66	89.5	92.5
HIGGS dataset	PSO	68.43	72.3	76.2
	BGOA	69.15	76.5	78.7
	HRBGOA	72.05	82.6	83.2

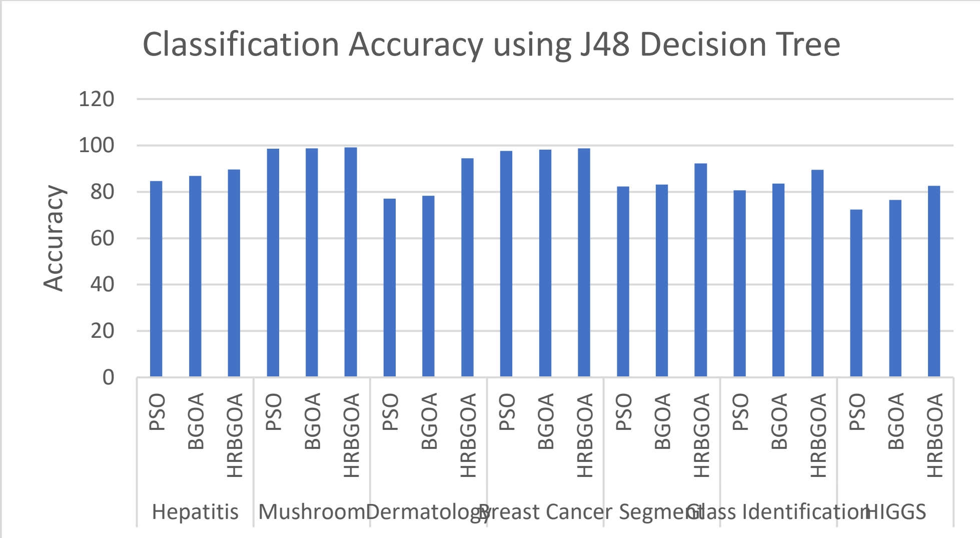

Figure 1: Classification Accuracy using J48 Decision Tree

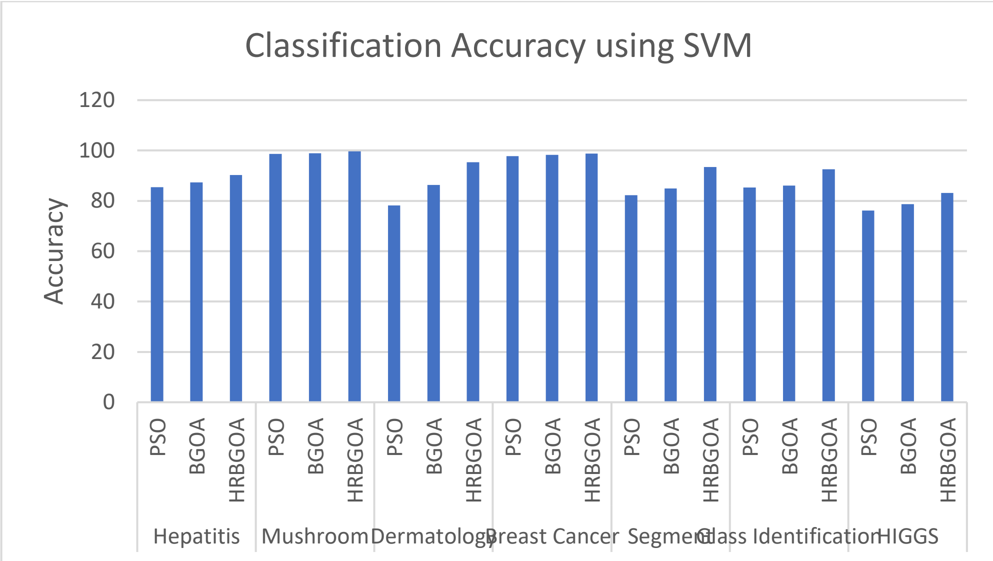

Figure 2: Classification Accuracy using SVM

In comparison to BGOA and PSO, HRBGOA produces superior results for seven datasets. Compared with other FS approaches, the proposed approach can provide better results. Contrasted with the work of Wah et al. [25], which used the breast cancer dataset, although the authors implemented different FS approaches; correlation, information gain, sequential forward selection, and sequential backward elimination, and achieved accuracy results of 79%, 93.5%, 97%, and 97%, respectively, our accuracy result was 98.8%, which is higher than others using the same dataset, indicating that we succeeded in selecting the optimal features and implementing appropriate machine learning techniques.

6.4 Analysis of Computational Time

The time required for running the algorithms on various data sets is illustrated in Table 3, in which various feature selection approaches with the J48 and SVM classifiers are applied. With each iteration, the goal function is executed 10 times. Table 3 shows how long each dataset takes to run HRBGOA, BGOA, or PSO. Objective function methods, such as determining the number of different values and the degree of reliance, took up most of their processing resources.

Table 3: Computational Time comparison of different approaches

Dataset	Approach	No. of Iterations	Time(S)
Hepatitis	PSO	20	0.9572
	BGOA	6	0.75
	HRBGOA	6	0.48
Mushroom	PSO	19	63.00201
	BGOA	14	50.48
	HRBGOA	3	19.39
Dermatology	PSO	20	5.51263
	BGOA	16	1.496
	HRBGOA	3	1.355
Breast Cancer	PSO	20	3.68599
	BGOA	6	1.682
	HRBGOA	6	1.4092
Segment	PSO	5	13.07107
	BGOA	6	12.78
	HRBGOA	4	5.010
Glass Identification	PSO	16	0.90
	BGOA	6	0.60
	HRBGOA	6	0.55
HIGGS dataset	PSO	20	95.44085
	BGOA	16	82.343
	HRBGOA	6	63.875

There are twenty iterations of the objective function. All datasets showed that HRBGOA took less time than BGOA despite the additional time needed for the new objective function that is employed in PSO.

7. Conclusion

This study projected HRBGOA as the classification filter FS technique for supplied datasets by utilizing BGOA with a Rough Set. It has produced a greater SR percentage while requiring fewer iterations, even while testing all features. Compared to BGOA and PSO techniques, the experimental findings reveal that HRBGOA produced improved FS in seven datasets. In comparison to the BGOA and PSO techniques, HRBGOA had a much higher SR percent and enhanced the J48 Decision Tree and SVM classification accuracy. Compared to the BGOA and PSO techniques, our methodology required fewer iterations to find the optimal answer. Experimental results have proved that BGOA with Rough Set has quicker computing time than BGOA with PSO.

In future work, we will integrate filter and wrapper approaches to examine the hybrid systems' performance. For instance, utilizing a filter approach for preparatory reduction of features, accompanied by the HRBGOA for fine-tuning. Additionally, to improve classification accuracy and handle complex datasets, we will work on incorporating the wrapper-based FS approach with deep learning models like Convolutional Neural Networks and Recurrent Neural Networks to be implemented in various domains, like economic or medical diagnostics.

References

[1] L. M. Abualigah, A. T. Khader, and E. S. Hanandeh, “A new feature selection method to improve the document clustering using particle swarm optimization algorithm,” Journal of Computational Science, vol. 25, pp. 456-466, 2018,doi:https://doi.org/10.1016/j.jocs.2017.07.018.

[2] S. H. Kafi, N. M. Mirza, A. J. Karam, A. Al-Saleh and A. Al-Zuky, "Automatic Numeral Recognition System Using Local Statistical and Geometrical Features," Iraqi Journal of Science, vol. 64, no. 4, pp. 1988-2000,Apr. 2023,doi:https://doi.org/10.24996/ijs.2023.64.4.35.

[3] J. Waleed, L. K. Almajmaie, A. N. Mazher and S. Albawi, "A Fast and Accurate Optimized Iris Recognition Scheme Based on a Modified GSO Algorithm," 2023 3rd International Scientific Conference of Engineering Sciences (ISCES), Diyala, Iraq, pp. 1-6, 2023,doi:https://doi.org/ 10.1109/ISCES58193.2023.10311496.

[4] H. Chen, W. Jiang, C. Li, and R. Li, “A heuristic feature selection approach for text categorization by using chaos optimization and genetic algorithm,” Mathematical Problems in Engineering, vol. 2013, pp. 1-6, 2013,doi:http://dx.doi.org/10.1155/2013/524017.

[5] S. S. Hameed, O. O. Petinrin, A. O. Hashi and F. Saeed, "Filter-wrapper combination and embedded feature selection for gene expression data," International Journal of Advances in Soft Computing and its Applications, vol.10, no. 1, pp. 90-105, 2018.

[6] V. F. Rodriguez-Galiano, J. A. Luque-Espinar, M. Chica-Olmo and M. P. Mendes, "Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods," Science of The Total Environment, vol. 624, pp. 661-672, 2018,doi:https://doi.org/10.1016/j.scitotenv.2017.12.152Get rights and content.

[7] S. H. Hammed and A. S. Albahri, "Unlocking the Potential of Autism Detection: Integrating Traditional Feature Selection and Machine Learning Techniques," Applied Data Science and Analysis, vol. 2023, pp. 42-58, 2023,doi:https://doi.org/10.58496/ADSA/2023/003.

[8] A. N. Mazher and J. Waleed, "Implementation of modified GSO based magic cube keys generation in cryptography," Eastern-European Journal of Enterprise Technologies, vol. 1, no. 9 (109), pp. 43-49, 2021,doi:https://doi.org/10.15587/1729-4061.2021.225508

[9] A. N. Mazher and J. Waleed, "Retina Based Glowworm Swarm Optimization for Random Cryptographic Key Generation," Baghdad Science Journal, vol. 19, no. 1, pp. 179-188, 2022,doi:https://doi.org/10.21123/bsj.2022.19.1.0179.

[10] Y. Chen, D. Miao, R. Wang, "A rough set approach to feature selection based on ant colony optimization," Pattern Recognition Letters, vol. 31, no. 3, pp. 226-233, 2010,doi:https://doi.org/10.1016/j.patrec.2009.10.013.

[11] S. Chebrolu and S. G. Sanjeevi, “Attribute reduction on real-valued data in rough set theory using hybrid artificial bee colony: extended FTSBPSD algorithm,” Soft Computing, vol. 21, no. 24, pp. 7543-7569, 2017,doi:https://doi.org/10.1007/s00500-016-2308-6.

[12] J. Waleed, H. D. Jun, T. Abbas, S. Hameed, and H. Hatem, "A Survey of Digital Image Watermarking Optimization based on Nature Inspired Algorithms NIAs," International Journal of Security and Its Applications, vol.8, no.6, pp. 315-334, 2014, doi: http://dx.doi.org/10.14257/ijsia.2014.8.6.28.

[13] J. Rajpurohit, T. K. Sharma, and A. Abraham, “Glossary of metaheuristic algorithms,” International Journal of Computer Information Systems & Industrial Management Applications, vol. 9, pp. 181-205, 2017,https://cspub-ijcisim.org/index.php/ijcisim/article/view/353.

[14] R. Eberhart and J. Kennedy, “A new optimizer using particle swarm theory,” in MHS’95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 2002,doi:https://doi.org/ 10.1109/MHS.1995.494215.

[15] M. Dorigo, V. Maniezzo and A. Colorni, "Ant system: optimization by a colony of cooperating agents," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 26, no. 1, pp. 29-41, 1996,doi:https://doi.org/10.1109/3477.484436.

[16] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey Wolf Optimizer," Advances in Engineering Software, vol. 69, pp. 46-61, 2014,doi:https://doi.org/10.1016/j.advengsoft.2013.12.007.

[17] S. Mirjalili, "The Ant Lion Optimizer," Advances in Engineering Software, vol. 83, pp. 80-98, 2015,doi:https://doi.org/10.1016/j.advengsoft.2015.01.010.
[18] S. Mirjalili, "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm," Knowledge-Based Systems, vol. 89, pp. 228-249, 2015, doi: https://doi.org/10.1016/j.knosys.2015.07.006.
[19] S. Mirjalili, A. Lewis, "The Whale Optimization Algorithm," Advances in Engineering Software, vol. 95, pp. 51-67, 2016, doi: https://doi.org/10.1016/j.advengsoft.2016.01.008.
[20] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," Applied Soft Computing, vol. 43, pp.117-130, 2016, doi: https://doi.org/10.1016/j.asoc.2016.01.044.
[21] Y. Zhang, H.-G. Li, Q. Wang and C. Peng, “A filter-based bare-bone particle swarm optimization algorithm for unsupervised feature selection,” Applied Intelligence, vol. 49, pp. 2889-2898, 2019, doi: https://doi.org/10.1007/s10489-019-01420-9.
[22] S. Gu, R. Cheng and Y. Jin, “Feature selection for high-dimensional classification using a competitive swarm optimizer,” Soft Computing, vol. 22, pp. 811-822, 2018, doi: https://doi.org/10.1007/s00500-016-2385-6.
[23] B. Xue, M. Zhang and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," Applied Soft Computing, vol. 18, pp. 261-276, 2014, doi: https://doi.org/10.1016/j.asoc.2013.09.018.
[24] M. A. Tawhid and K. B. Dsouza, “Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems,” Applied Computing and Informatics, vol. 16, no. 1/2, pp. 117-136, 2018,https://www.emerald.com/insight/2210-8327.htm.
[25] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman and S. Fong, "Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy, " Pertanika Journal of Science and Technology, vol. 26, no. 1, pp. 329-340, 2018.
[26] Z. Pawlak, "Rough sets", International Journal of Computer & Information Sciences, vol. 11, no. 5, pp. 341-356, 1982, doi: https://doi.org/10.1007/BF01001956.
[27] S. Q. Wang, C. Y. Gao, C. Luo, G. M. Zheng, Y. N. Zhou, "Research on Feature Selection/Attribute Reduction Method Based on Rough Set Theory," Procedia Computer Science, vol. 154, pp. 194-198, 2019, doi: https://doi.org/10.1016/j.procs.2019.06.029.
[28] P. R. K. Varma, V. V. Kumari, and S. S. Kumar, “A novel rough set attribute reduction based on ant colony optimisation,” International Journal of Intelligent Systems Technologies and Applications, vol. 14, no. 3/4, p. 330 - 353, 2015, doi: https://doi.org/10.1504/IJISTA.2015.074333.
[29] W. A. Mahmood, L. K. Almajmaie, A. R. Raheem, S. Albawi "A hybrid approach towards movie recommendation system with collaborative filtering and association rule mining," Acta Scientiarum -Technology, vol.44, e58925, 2022, doi: https://doi.org/ 10.4025/actascitechnol.v44i1.58925.
[30] L. K. Almajmaie, A. R. Raheem, W. A. Mahmood, S. Albawi, "MRI image segmentation using machine learning networks and level set approaches," International Journal of Electrical and Computer Engineering, vol.12910, pp. 793-801, 2022, doi: https://doi.org/10.11591/ijece.v12i1.pp793-801.
[31] Z.T. Abbood, M. N. H. Al-Turfi, L. K. Almajmaie "An abbreviated review of deep learning-based image classification models," Indonesian Journal of Electrical Engineering and Computer Science, vol.30, no. 1, pp. 491-500, 2023, doi: http://doi.org/10.11591/ijeecs.v30.i1.pp491-500.