Iran Journal of Computer Science
https://doi.org/10.1007/s42044-024-00174-z

RESEARCH

An efficient high-dimensional gene selection approach based on the Binary Horse Herd Optimization Algorithm for biologicaldata classification

Niloufar Mehrabi[1] · Sayed Pedram Haeri Boroujeni[1] · Elnaz Pashaei[2]

Received: 30 November 2023 / Accepted: 19 January 2024

Abstract
The Horse Herd Optimization Algorithm (HOA) is a new meta-heuristic algorithm inspired by the behaviors of horses of different ages. It was recently introduced to solve complex and high-dimensional problems. This paper proposes a binary version of the HOA, termed the Binary Horse Herd Optimization Algorithm (BHOA), designed to solve discrete problems and select prominent feature subsets. Additionally, this study introduces a novel hybrid feature selection framework that combines the BHOA with a Minimum Redundancy Maximum Relevance (MRMR) filter method. This hybrid approach, more computationally efficient, yields a beneficial subset of relevant and informative features. Recognizing feature selection as a binary problem, we have applied a new Transfer Function (TF), named the X-shape TF, which converts continuous problems into binary search spaces. Moreover, the Support Vector Machine (SVM) is employed to evaluate the efficiency of the proposed method on ten microarray datasets: Lymphoma, Prostate, Brain-1, DLBCL, SRBCT, Leukemia, Ovarian, Colon, Lung, and MLL. Compared to other state-of-the-art methods, such as the Gray Wolf (GW), Particle Swarm Optimization (PSO), and Genetic Algorithm (GA), our proposed hybrid method (MRMR-BHOA) demonstrates superior performance in terms of accuracy and minimal selected features. Experimental results also show that the X-shaped BHOA approach outperforms other methods.

Keywords Feature selection · Support Vector Machine (SVM) · Transfer function · Swarm intelligence algorithm

1 Introduction

In cancer research, microarray technology has become instrumental, enabling researchers to simultaneously assess the expression of thousands of genes [1, 2]. This approach is vital in discovering genes that serve as biomarkers, which are critical for diagnostic, prognostic, and therapeutic purposes. This technology significantly applies to differentiating between normal and cancerous tissues. However, analyzing microarray data is challenging due to the high number of genes compared to the relatively small number of samples, making it difficult to identify relevant biomarkers and classify tissue types accurately. Additionally, the data's complexity is heightened by the interrelation of genes, many of which may be redundant or irrelevant for clinical applications. Gene selection is, therefore, a critical step in bioinformatics, especially with high throughput data like microarrays [3]. These datasets are characterized by a vast number of genes, most of which are unnecessary or repetitive, overshadowing the few genuinely informative ones. Removing these irrelevant and redundant genes is crucial for improving the accuracy of classification models. The primary aim of gene selection is to identify a small subset of genes with significant discriminative ability, thereby enhancing the predictive performance of these models. This process not only reduces computational demands by decreasing data size but also helps in making more precise predictions. Effective gene selection or Feature Selection (FS) is thus vital in aiding clinicians in making

✉ Niloufar Mehrabi
nmehrab@g.clemson.edu

Sayed Pedram Haeri Boroujeni
shaerib@g.clemson.edu

Elnaz Pashaei
epashaei@iu.edu

[1] School of Computing, Clemson University, Clemson, SC, USA

[2] Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, USA

Published online: 08 February 2024

accurate diagnoses and choosing the most appropriate treatment options.

Feature Selection should be applied in machine learning as a pre-processing phase to get optimal output with short training times and low memory consumption [4]. FS plays a significant role in data mining [5] to solve various problems, such as data classification [6], data clustering [7], image processing [8], text clustering [9], disaster management [10], and disease forecasting [11]. FS is generally classified into three major groups based on a variety of evaluation criteria, i.e., filter method [12], wrapper model [13], and embedded technique [14].

Filter methods measure the relevance of features by their correlation with the dependent variable. Also, this technique uses statistical methods to evaluate a subset of features [15]. Filters select some features solely based on intrinsic characteristics of the training data without involving any learning algorithm. As a result, they are significantly efficient, easily adaptable to more complex microarray datasets, and are readily applied using only a single application to produce results. Nevertheless, it tends to choose subsets with many features (even all of them), so a threshold is needed to select a subset. There are a variety of filter methods, such as Relief [16], Information Gain [17], and Chi-square [18].

On the other hand, wrapper methods measure the usefulness of a subset of features by actually training a model on it [19]. However, wrapper methods are also computationally very costly. Wrapper algorithms can evaluate interactions between genes based on the classifier's accuracy. The wrapper methods achieve a higher rate of predictive accuracy than filter methods. Also, the embedded technique in feature selection is an integral approach where feature selection is performed as part of the model training process. By embedding the feature selection within the learning algorithm, this method ensures that the chosen features are optimized for that specific model. It offers the advantage of being more efficient and model-specific than standalone feature selection methods, as it simultaneously learns the pattern and selects the most relevant features [20].

All three feature selection methods exhibit their drawbacks, necessitating careful consideration of their application:

1. Filter methods, while assessing features based on intrinsic data characteristics, often need to pay more attention to the context of the learning model, potentially leading to suboptimal selections.
2. Their tendency to oversimplify complex data interrelations can result in the inclusion of redundant features. Conversely, although tailored to enhance model performance, wrapper methods incur significant computational demands due to the necessity of training a model for each feature subset. This increases the risk of overfitting the training data and poses scalability challenges, particularly in datasets with a vast number of features.
3. Embedded methods are intricately tied to specific machine learning models, limiting their applicability across different models.
4. Their inherent complexity complicates interpretability, and the integration of feature selection within the learning process reduces flexibility.

Thus, selecting an appropriate method should be guided by a balanced consideration of these trade-offs, aligning with the specific requirements and constraints of the task.

The feature selection is defined as an NP-hard problem. For a given microarray dataset with N number of genes, $2n$ possible subsets must be evaluated to demonstrate the best gene subsets. Due to the search space expanding exponentially as the number of features increases, it is impossible to conduct an exhaustive search for the optimal feature subset in a high-dimensional space [21].

Bioinformatics relies heavily on feature selection to reduce dimensionality and improve model performance. Traditional methods like filtering, wrapping, and embedding techniques have been extensively used but often need to improve in handling microarray data's high dimensionality and complexity. Recent literature emphasizes the growing application of meta-heuristic algorithms in feature selection, offering solutions where conventional methods lag. For instance, Improved Whale Optimization and Chimp Optimization Algorithms have shown promising results in optimization and feature selection tasks.

Meta-heuristic algorithms (MOAs), particularly binary variants used in feature selection, exhibit several limitations that merit careful consideration. A primary drawback is their potential for premature convergence, where the algorithm might settle on a local optimum rather than finding the most optimal global solution. This issue is particularly pronounced in complex search spaces, where the balance between exploring new solutions and exploiting known good ones is crucial. Furthermore, these algorithms require careful calibration of parameters, which can be a meticulous and time-consuming process, often necessitating trial and error to achieve desirable outcomes. In addition, binary meta-heuristic algorithms can incur substantial computational costs, especially in scenarios involving large datasets, due to their iterative nature and the need to evaluate numerous possible solutions. Another challenge is their inconsistency in performance across different datasets or problems, as these algorithms may excel in some scenarios but underperform in others. Consequently, while binary meta-heuristic algorithms are potent tools for feature selection, their application demands a thoughtful approach, considering these limitations to leverage their capabilities thoroughly.

Swarm Intelligence (SI) is a subset of meta-heuristic algorithms inspired by the behavior of birds, wolves, dragonflies, whales, and other animals. Improved Whale Optimization Algorithm (WOA) [22], Chimp Optimization Algorithm (ChOA) [23], Flying Squirrel Optimizer (FSO) [24], Coyote Optimization Algorithm (COA) [25], Harris Hawks Optimization (HHO) [26], Grasshopper Optimization Algorithm (GOA) [27], and Intelligent Dynamic Genetic Algorithm (IDGA) [28] are some recent works of SI that utilized for solving optimization problems.

SI algorithms excel in feature selection due to their robustness and adaptability, effectively handling high-dimensional datasets. They adeptly balance exploration and exploitation, which is crucial for identifying optimal features, and their parallel processing capabilities significantly speed up the selection process. Furthermore, these algorithms reduce the risk of overfitting and do not rely on gradient information, making them versatile and straightforward to implement. Consequently, their intuitive nature and efficiency make them an invaluable asset in feature selection.

MOAs' success in dealing with various optimization cases eventually led us to integrate SIs into feature selection problems. Based on this idea, we introduce the binary Horse Herd Optimization Algorithm (BHOA) as a response to the limitations and drawbacks of other swarm intelligence algorithms.

In addressing the challenges of feature selection in gene analysis, our research introduces the Binary Horse Herd Optimization Algorithm (BHOA), a novel adaptation of the Horse Herd Optimization Algorithm (HOA) that operates in a binary context. The traditional HOA [29], a recent meta-heuristic approach based on swarm intelligence, mimics various behavioral patterns of horse herds, such as Grazing, Hierarchy, Sociability, Imitation, Defense mechanism, and Roaming. These behaviors are instrumental in balancing exploration and exploitation phases in solving simple and complex optimization problems, especially those involving high-dimensional data. The BHOA specifically targets the discrete nature of feature selection, effectively mapping the continuous search space of HOA into a binary one. This transformation is facilitated by a novel X-shaped transfer function [33], which has demonstrated superior performance compared to the more traditional S-shaped and V-shaped transfer functions. This advancement represents a direct response to the limitations of existing binary meta-heuristic algorithms in feature selection, building upon recent advancements in this field.

Additionally, our approach incorporates the Minimum Redundancy Maximum Relevance (MRMR) method as a filter technique in the initial phase of the hybrid gene selection process, effectively reducing redundant and irrelevant genes in gene expression datasets [34]. In machine learning, the effectiveness of models is determined by their precision in making predictions on new, unseen datasets, which is crucial for confirming their usability and dependability in various real-life applications [30].

The efficacy of BHOA is rigorously evaluated using two classifiers, Naïve Bayes (NB) and Support Vector Machine (SVM), across ten well-known datasets, employing a ten-fold cross-validation method to establish the effectiveness of each classifier. Our findings indicate that SVM, which performs optimally in high-dimensional spaces and when the number of dimensions exceeds the number of samples, is an ideal evaluator for the selected genes in our proposed hybrid approach. Finally, the performance of BHOA is compared with existing nature-inspired optimization methods like Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Grey Wolf Optimizer (GWO) through statistical and convergence rate analyses, demonstrating its efficacy in the context of feature selection in bioinformatics.

The critical difference between our proposed method and other existing algorithms lies in its unique Binary Horse Herd Optimization Algorithm (BHOA), which explicitly targets feature selection in gene analysis. This method efficiently converts continuous problems into binary ones using a novel X-shaped Transfer Function. A significant aspect of our method is integrating the MRMR (Minimum Redundancy Maximum Relevance) filter with the BHOA, creating a hybrid solution that enhances feature selection effectiveness. This integration suggests a more robust and effective feature selection process, as it combines the strengths of both techniques. Moreover, a noteworthy advantage of our approach is its capability to avoid local minima, ensuring a more thorough and accurate exploration of the solution space. This characteristic is particularly crucial in the intricate field of gene analysis, where local minima often hinder finding the optimal solution. Our method not only simplifies the complexity inherent in gene analysis but also provides a more reliable and efficient pathway to feature selection. Additionally, our algorithm excels in rapidly finding optimal solutions and demonstrates superior accuracy and minimal feature selection across various datasets, distinguishing it from other existing methods.

Therefore, the main contributions of this work can be summarized as follows:

- We introduce a binary variant of the Horse Herd Optimization Algorithm (HOA), tailored explicitly to bioinformatics's gene selection problem. This adaptation represents a novel approach in the field, addressing the unique challenges of selecting genes from complex datasets.
- Our work involves an extensive examination of nine different transfer functions, including X-shaped, V-shaped, and S-shaped functions, in the context of the Horse Herd Optimization Algorithm. This study is crucial for understanding how these transfer functions influence the

algorithm's ability to convert continuous search spaces into binary formats, an essential step in effective gene selection.
- A significant focus of our research is developing a hybrid gene selection method. This method combines the MRMR (Minimum Redundancy Maximum Relevance) filter approach with the newly developed binary HOA. The integration of these two approaches represents an innovative strategy in the field, aiming to enhance the accuracy and efficiency of gene selection.
- The performance of the proposed method is evaluated using the Support Vector Machine (SVM) on ten microarray datasets, namely Lymphoma, Prostate, Brain-1, DLBCL, SRBCT, Leukemia, Ovarian, Colon, Lung, and MLL. Moreover, the results are compared to other state-of-the-art, such as the Gray Wolf (GW), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Firefly, and ACO.
- Ultimately, our method demonstrates a higher capability for addressing feature selection issues, achieving a notable enhancement in securing the least number of features while maximizing accuracy, outperforming other techniques across all datasets evaluated. Additionally, our proposed algorithm exhibits a quicker convergence rate, rapidly attaining the optimal solution in the early stages of iterations.

The structure of this study is outlined as follows for ease of understanding: Sect. 2 provides a brief overview of relevant prior research. Next, Sects. 3 and 4 thoroughly describe our framework, which includes the continuous Horse Herd Optimization Algorithm (HOA), its binary version, the MRMR filter method, the X-shaped transfer function, and our newly developed MRMR-BHOA hybrid approach for feature selection. Section 5 details our experimental procedures, presents the results, compares our approach to others, and discusses these findings. Finally, the last section summarizes our conclusions.

2 Literature review

In the field of feature selection, meta-heuristic algorithms have gained prominence due to their adaptability and efficiency in navigating complex optimization challenges. These algorithms are generally categorized into three main types: Evolutionary Algorithms (EA), Physics-based algorithms, and Swarm Intelligence (SI) algorithms. Each category, with its unique methodologies and principles, contributes significantly to advancements in feature selection, particularly in gene analysis.

The first category of meta-heuristic algorithms, Evolutionary Algorithms (EA), draws inspiration from biological evolution and the natural laws articulated by Darwin. These algorithms, including the widely recognized Genetic Algorithm (GA), are grounded in the principles of natural selection and survival of the fittest [35]. GA, for instance, applies these concepts to algorithmic problem-solving by simulating the process of natural selection, where the most adaptable solutions survive, reproduce, and evolve over successive generations. This approach effectively enables GAs to tackle complex, non-linear problems and multi-objective optimization challenges. EAs, despite their flexibility and depth, have limitations. A significant challenge is their tendency to converge prematurely on local minima, potentially overlooking more optimal global solutions. This issue is particularly acute in high-dimensional and rugged search landscapes. Other notable examples within this category, such as Simulated Annealing (SA) [36], Genetic programming (GP) [37], Memetic Algorithm (MA) [38], and Gradient Evolution Algorithm (GEA) [39], each bring their unique strengths and face similar challenges, including balancing exploration and exploitation, maintaining diversity in solutions, and computational efficiency. These limitations underscore the need for continued innovation in EA methodologies to enhance their applicability and effectiveness in complex problem-solving scenarios.

Physics-based meta-heuristic algorithms represent another significant category in this domain, distinguished by their grounding in fundamental physical laws and phenomena, such as electromagnetic forces, inertia, and gravitational principles [40]. This category includes notable algorithms such as the Vibrating Particles System (VPS) [41], Binary Multi-Verse Optimizer (MVO) [42], Binary Dragonfly Algorithm (BDA) [43], Binary Gravitational Search Algorithm (GSA) [44], and Ideal Gas Molecular Movement (IGMM) [45]. Each of these algorithms employs physical concepts to structure their search and optimization processes. For instance, the Binary Multi-Verse Optimizer (BMVO) is a unique model incorporating cosmological theories such as wormholes, black holes, and white holes, providing a mathematical framework for exploring and exploiting the search space. In BMVO, solutions are represented in binary form, and a V-shaped function is used to transition continuous values into this binary format. Notably, BMVO is characterized by its rapid convergence speed.

Another prominent algorithm in this category is the Binary Gravitational Search Algorithm (BGSA), which leverages Newton's law of gravitation. This algorithm simulates mass interactions to guide its search for optimal solutions, effectively avoiding local optima through stochastic rules. In BGSA, the 'heaviest mass' concept denotes the optimal solution, as it is the most influential in attracting other elements within the search space [46]. The algorithm translates gravitational forces into probabilities, determining the binary states of elements (0 or 1).

Despite their innovative approaches and efficiency in specific contexts, physics-based algorithms also encounter limitations. Their reliance on physical principles might restrict adaptability in specific problem spaces, particularly where the abstraction from physical to algorithmic models may need to align with the problem's nature perfectly. Additionally, the translation of continuous values into binary forms, while effective, may introduce complexities or inaccuracies in representing certain types of data or search spaces. As such, while these algorithms offer unique perspectives and methods for optimization, their application must be carefully considered in the context of the specific problems and data sets they are employed to solve.

Swarm Intelligence (SI) represents the final subcategory in meta-heuristic optimization algorithms, characterized by its emulation of collective animal behaviors observed in species such as birds, bats, horses, wolves, and chimpanzees [47]. Key algorithms in this category include Binary Particle Swarm Optimization (BPSO) [48], Binary Bat Algorithm (BBA) [49], Binary Whale Optimization Algorithm (BWOA) [50], Flying Squirrel Optimizer (FSO) [24], Binary Harris Hawks Optimization (BHHO) [51], Binary Emperor Penguin Optimizer (BEPO) [52], Binary Dragonfly Algorithm (BDA) [43], Improved Binary Grey Wolf Optimization (IBGWO) [53], and Binary Coyote Optimization (COA) [20].

The foundational Particle Swarm Optimization (PSO) algorithm, developed by Eberhard and Kennedy in 1995, is based on animal social behaviors, particularly collaborative foraging strategies. PSO is renowned for its rapid search and optimization capabilities within a multidimensional space, starting with a randomly selected population of solutions, termed 'particles' [54]. Each particle in the population represents a potential solution and navigates the search space by updating its position based on its velocity and the most favorable position discovered so far. This algorithm is lauded for its ability to swiftly converge to high-quality solutions, making it practical for a variety of optimization problems [55].

The Chimp Optimization Algorithm (ChOA) exemplifies the ingenuity of Swarm Intelligence algorithms by emulating chimpanzees' complex social and hunting behaviors [56]. It integrates chimps' diverse roles and strategies—drivers, barriers, chasers, and attackers—each essential in different hunting phases. These roles range from pursuing prey to blocking escape routes, with effectiveness depending on factors like age and intelligence. The algorithm's success hinges on accurately balancing exploration (such as driving and chasing) with exploitation (like attacking), a critical aspect of optimization.

The Emperor Penguin Optimizer (EPO), another SI-inspired algorithm, draws from the huddling behavior of emperor penguins [57]. It involves steps like defining huddle boundaries and calculating temperatures, which are crucial for determining efficient movement within the group. However, its inability to handle discrete and binary problems led to the development of the Binary Emperor Penguin Optimizer (BEPO) [52], which adapts the penguins' spatial dynamics into binary search spaces using S-shaped and V-shaped transfer functions. BEPO has been effectively utilized in feature selection challenges.

Lastly, the Coyote Optimization Algorithm (COA), inspired by the behavior of Canis latrans, reflects a recent advancement in this domain [25]. COA considers coyotes' social structure and environmental adaptability, introducing innovative mechanisms to balance exploration and exploitation in optimization processes. It was initially developed for global optimization with continuous values based on simple parameters like pack size and number of coyotes per pack. The binary variant, Binary Coyote Optimization Algorithm (BCOA), extends this methodology to binary search spaces, showcasing remarkable precision and convergence capabilities in feature selection tasks [20].

While Swarm Intelligence (SI) algorithms, inspired by animal behavior, offer innovative optimization approaches, they have inherent challenges. These algorithms often need help accurately translating intricate animal behaviors into effective algorithmic solutions, particularly for abstract or non-intuitive problems. Moreover, their performance varies widely across problem domains and search spaces, sometimes necessitating specialized adaptations, such as binary versions for discrete problem-solving. This inconsistency underscores a significant gap in the current optimization landscape. Recognizing these limitations motivates the development of our new algorithm. We aim to provide a more adaptable and universally effective solution that consolidates SI algorithms' strengths while addressing their weaknesses. This algorithm is designed to perform consistently across diverse problem types and search spaces, offering a more reliable and efficient tool for complex optimization tasks.

3 Continuous Horse Herd Optimization Algorithm

The Horse Herd Optimization Algorithm was developed based on the behavior of horses in their natural environment at various ages [29]. The Horse Herd Optimization Algorithm (HOA) is designed to replicate the complex social structures and behaviors of horse herds in nature. It specifically models how horses interact and move in a herd, reflecting their social hierarchy, grazing behavior, and group dynamics. The algorithm applies these behaviors to optimization problems by simulating how different horses, categorized by age and role, would explore (search for new areas) and exploit (make the best out of known areas) in their environment. This modeling is aimed at effectively navigating the solution space of

a problem, ensuring a thorough exploration to avoid local optima and improve the chances of finding the global optimum. The algorithm's structure mirrors the adaptive and efficient nature of horse herds, applying these qualities to solve complex optimization problems. Horse behavior can be divided into six categories: Grazing, Hierarchy, Sociability, Imitation, Defense Mechanism, and Roaming. Every horse moves according to the following equation during each iteration:

$$X_i^{\text{iter, AGE}} = V_i^{\text{iter, AGE}} + X_i^{(\text{iter}-1),\text{AGE}}, \quad \text{AGE} = \alpha,\ \beta,\ \gamma,\ \delta \tag{1}$$

where, $X_i^{\text{iter, AGE}}$ demonstrates the location of the ith horse, $V_i^{\text{iter, AGE}}$. presents the velocity of the i_{th} horse, AGE indicates the age range of each horse, and iter shows the current iteration. Horses' behavior changes at different ages. Each horse has a lifespan of 25–30 years on average. Regarding this, δ. refers to the ages between 0 and 5 years, γ. denotes the horses between the ages 5 and 10, β represents the age ranges from 10 to 15 years, and α shows the horses older than 15 years. The horse ages should be selected based on a matrix of responses generated per iteration. Thus, the matrix can be arranged depending on which answers are the most effective, and according to Fig. 1, α horses are those horses ranked in the top 10% of the sorted matrix. The ubsequent 20% belong to the β group. Regarding the remaining horses, the γ and δ horses are considered 30% and 40%, respectively. The velocity of each horse can be calculated based on its age group and behavior patterns within an iteration.

$$\begin{aligned}
V_i^{\text{iter},\alpha} &= G_i^{\text{iter},\alpha} + D_i^{\text{iter},\alpha} \\
V_i^{\text{iter},\beta} &= G_i^{\text{iter},\beta} + H_i^{\text{iter},\beta} + S_i^{\text{iter},\beta} + D_i^{\text{iter},\beta} \\
V_i^{\text{iter},\gamma} &= G_i^{\text{iter},\gamma} + H_i^{\text{iter},\gamma} + S_i^{\text{iter},\gamma} + I_i^{\text{iter},\gamma} \\
&\quad + D_i^{\text{iter},\gamma} + R_i^{\text{iter},\gamma} \\
V_i^{\text{iter},\delta} &= G_i^{\text{iter},\delta} + I_i^{\text{iter},\delta} + R_i^{\text{iter},\delta}
\end{aligned} \tag{2}$$

Algorithm 1: Horse Herd Optimization Algorithm

```
Define input parameters
Generate random positions for n horses
Evaluate the fitness of each horse's location
Generate Global Matrix based on the horses' location and their fitness value
    while (the stopping criterion is not satisfied) do
        for i = 1: total number of horses do
            Sort the locations of horses in ascending order depending upon their fitness value
            Calculate the mean position by Equation (9)
            Calculate the good position by Equation (12)
            Calculate the bad position by Equation (15)
            Determining alpha, beta, gamma, and delta horses
            Computing the velocity of each horse by Equation (2)
            Update the position of each horse using Equation (1)
            Evaluate fitness for the new position of each horse
            if new fitness value < old fitness value, then
                set the new position as the best position
                set new fitness value as the best fitness value
            end if
        end for
    end while
return the best position
return the best fitness value
```

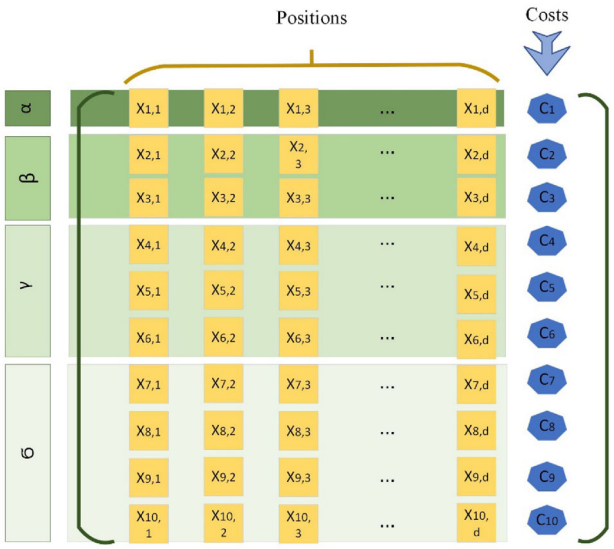

Fig. 1 The Sorted Matrix of responses in the HOA algorithm

3.1 Grazing (G)

Grazing is one of the most common horse behavior patterns. Horses graze roughly 70% of the time during the day and 50% at night. In other words, at any age, they spend approximately 16–20 h per day grazing on pastures. The HOA algorithm mods the grazing area around each horse using the coefficient g. Equations (3) and (4) describe grazing mathematically.

$$G_i^{\text{iter, AGE}} = g_{\text{iter}}\left(\breve{u} + \mathcal{P}\breve{l}\right)\left[X_i^{(\text{iter}-1)}\right], \quad \text{AGE} = \alpha, \beta, \gamma, \delta \tag{3}$$

$$g_i^{\text{iter, AGE}} = g_i^{(\text{iter}-1),\,\text{AGE}} \times \omega_{\text{g}} \tag{4}$$

where $G_i^{\text{iter, AGE}}$ measures the tendency of the ith horse to graze that decreases linearly by ω_{g} with each iteration. In addition, $\breve{l}$ and $\breve{u}$ indicate the lower and upper boundary of the grazing area, respectively, while $\mathcal{P}$ is a random value between 0 and 1. For all age groups, it is suggested [29] to assign 0.95 and 1.05 for $\breve{l}$ and $\breve{u}$, respectively, and the coefficient g is 1.5.

3.2 Hierarchy (H)

There is always a hierarchy among horses because they are herd animals. The hierarchy protects horses while also allowing them access to better feeding grounds. In the HOA algorithm, the coefficient h represents the tendency of horses to follow the horse with the most experience and strength. Many investigations have shown that in the Middle Ages, β and γ, horses tended to have hierarchical behavior [31]. Equations (5) and (6) describe hierarchy behavior as follows:

$$H_i^{\text{iter, AGE}} = h_i^{\text{iter, AGE}}\left[X_*^{(\text{iter}-1)} - X_i^{(\text{iter - 1})}\right],$$
$$\text{AGE} = \alpha, \beta, \gamma \tag{5}$$

$$h_i^{\text{iter, AGE}} = h_i^{(\text{iter - 1}),\,\text{AGE}} \times \omega_{\text{h}} \tag{6}$$

where $X_*^{(\text{iter}-1)}$ demonstrates the bestorse's location, and $H_i^{\text{iter, AGE}}$ shows how the best horse's location affects the velocity parameter.

3.3 Sociability (S)

Social behavior is a characteristic of horses [32]. The purpose of sociability is to increase predator protection mechanisms and decrease scanning time. Furthermore, sociability increases intra-group competition and conflicts, disease transmission, and the risk of attracting predators. Since flight is horses' top protection mechanism, it is vital to identify hidden predators immediately. Therefore, maintaining shared awareness, acting as a group during flight, and communicating effectively about such actions should be prime importance for each herd horse. Observation shows that horses aged 5–15 are very interested in social life in a group. Factor s demonstrates the social behavior of horses, which is described by Eqs. (7) and (8).

$$S_i^{\text{iter, AGE}} = s_i^{\text{iter, AGE}}\left[\left(\frac{1}{N}\sum_{j=1}^{N} X_j^{(\text{iter}-1)}\right) - X_i^{(\text{iter - 1})}\right],$$
$$\text{AGE} = \beta, \gamma \tag{7}$$

$$s_i^{\text{iter, AGE}} = s_i^{(\text{iter}-1),\,\text{AGE}} \times \omega_{\text{s}} \tag{8}$$

$S_i^{\text{iter, AGE}}$ denotes the social movement vector of the ith horse, decreasing per iteration by the ω_{s} factor. Also, N is the total number of horses. AGE expresses the age group of horses. Furthermore, the average position is obtained as follows:

$$\text{Mean\_Position} = \left(\frac{1}{N}\sum_{j=1}^{N} X_j^{(\text{iter}-1)}\right) \tag{9}$$

3.4 Imitation (I)

As social animals, horses can learn from each other about good and bad behavior, such as discovering their suitable grassland [29]. Young horses tend to imitate other horses in the age range of 0–5 years. In addition, factor i in the HOA algorithm exhibits this feature of horses' behavior. Imitation in horse herds can be defined as follows:

$$I_i^{\text{iter, AGE}} = i_i^{\text{iter, AGE}} \left[\left(\frac{1}{\text{pN}} \sum_{j=1}^{\text{pN}} \hat{X}_j^{(\text{iter}-1)} \right) - X^{(\text{iter}-1)} \right],$$
$$\text{AGE} = \gamma \tag{10}$$

$$i_i^{\text{iter, AGE}} = i_i^{(\text{iter}-1),\ \text{AGE}} \times \omega_{\text{i}} \tag{11}$$

$I_i^{\text{iter, AGE}}$ demonstrates the movement of the ith horse toward the mean position of the best horses, which are located at $\hat{X}$. pN indicates the number of horses in the best position. The recommended value of p is 10% of the total horses. Also, ω_{i} can be expressed as a reduction factor for each iteration, as mentioned previously. Moreover, the best location is obtained by the following equation.

$$\text{Best\_Position} = \left(\frac{1}{\text{pN}} \sum_{j=1}^{\text{pN}} \hat{X}_j^{(\text{iter}-1)} \right) \tag{12}$$

3.5 Defense mechanism (D)

Horses have two primary defense mechanisms: flight and fight. When confronted with a dangerous situation, horses typically flee, and fighting is a secondary survival mechanism. Factor d describes the defense system of horses. Equations (13) and (14) with negative coefficients represent the horse's protective mechanisms, which prevent them from getting into inappropriate positions.

$$D_i^{\text{iter, AGE}} = -d_i^{\text{iter, AGE}} \left[\left(\frac{1}{\text{qN}} \sum_{j=1}^{\text{qN}} \hat{X}_j^{(\text{iter}-1)} \right) - X^{(\text{iter}-1)} \right],$$
$$\text{AGE} = \alpha,\ \beta,\ \gamma \tag{13}$$

$$d_i^{\text{iter, AGE}} = d_i^{(\text{iter}-1),\ \text{AGE}} \times \omega_{\text{d}} \tag{14}$$

$D_i^{\text{iter, AGE}}$ denotes the fleeing vector of the ith horse based on the mean of horses with the worst areas. The number of horses with the worst locations is also shown by qN. It is proposed that q corresponds to 20 percent of all horses. As previously stated, ω_{d} represents the reduction factor for each iteration. Furthermore, the worst position is calculated as follows:

$$\text{Worst\_Position} = \left(\frac{1}{\text{qN}} \sum_{j=1}^{\text{qN}} \hat{X}_j^{(\text{iter}-1)} \right) \tag{15}$$

3.6 Roaming (R)

At the age range of 5–15 years, younger horses are more likely to move from pasture to pasture in looking for food. Horses are curious animals who seek out new pastures wherever they can. As horses reach maturity, they lessen their roaming behavior. The factor r shows this behavior of horses as a random motion in Eqs. (16) and (17).

$$R_i^{\text{iter, AGE}} = r_i^{\text{iter, AGE}} P X^{(\text{iter}-1)}, \quad \text{AGE} = \gamma,\ \delta \tag{16}$$

$$r_i^{\text{iter, AGE}} = r_i^{(\text{iter}-1),\ \text{AGE}} \times \omega_{\text{r}} \tag{17}$$

In this case, ith horse's random velocity vector is represented by $R_i^{\text{iter, AGE}}$, and its reduction factor is indicated by ω_{r}. To calculate the general velocity, the grazing, sociability, hierarchy, defense mechanism, imitation, and roaming are substituted into Eq. (2).

By applying a sorting mechanism in a global matrix, HOA uses an appropriate technique to boost the speed of issue resolution while also avoiding local optimal entrapment. The global matrix in this study represents a crucial component in the Horse Herd Optimization Algorithm (HOA). It is designed to organize and evaluate potential solutions systematically based on specific criteria. The intuition behind this matrix is to create a structured way of comparing and selecting the most promising solutions from a pool of possibilities. By juxtaposing positions and the cost of each position, the global matrix facilitates a transparent and efficient process for identifying the best solutions, aiding in the optimization process. This approach is particularly effective in managing and navigating through the complex solution spaces typical in optimization problems. The pseudo-code of HOA is stated in Algorithm 1. As indicated in Eqs. (18) and (19), the global matrix is constructed by juxtaposing positions (X) and the cost value for every position ($C(X)$).

$$X = \begin{bmatrix} x_{1,1} x_{1,2} x_{1,3} \cdots x_{1,d} \\ x_{2,1} x_{2,2} x_{2,3} \cdots x_{2,d} \\ \vdots \vdots \ddots \vdots \\ x_{m,1} x_{m,2} x_{m,3} \cdots x_{m,d} \end{bmatrix}, \quad C(X) = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} \tag{18}$$

$$\text{Global Matrix} = [XC(X)] = \begin{bmatrix} x_{1,1} x_{1,2} x_{1,3} \cdots x_{1,d} c_1 \\ x_{2,1} x_{2,2} x_{2,3} \cdots x_{2,d} c_2 \\ \vdots \vdots \ddots \vdots \\ x_{m,1} x_{m,2} x_{m,3} \cdots x_{m,d} c_m \end{bmatrix} \tag{19}$$

4 Proposed approach

4.1 Minimum Redundancy Maximum Relevancy (MRMR)

The Minimum Redundancy Maximum Relevance (MRMR) method is used for feature selection in machine learning and data analysis [58]. The intuition behind MRMR is to find a subset of features highly relevant to the target variable (maximum relevance) while being minimally redundant with each other (minimum redundancy). The MRMR method uses specific criteria and equations to measure the relevance of each feature with the target variable and the redundancy among the features. By balancing these two aspects, MRMR effectively selects features that contribute the most to predicting the target variable while reducing the complexity of the model. This method is beneficial in scenarios where reducing the number of features without losing predictive power is crucial. This technique basically identifies the features with maximum relevancy and minimum redundancy [34]. Mutual information (MI) quantifies both relevancy and redundancy by measuring the mutual dependence of two variables. A definition of mutual information is given by Eq. (20).

$$\mathrm{MI}(x_i, C) = \sum_{x \in X} \sum_{y \in Y} P(x_i, C) \log \frac{P(x_i, C)}{P(x_i) P(C)} \tag{20}$$

where, $\mathrm{MI}(x_i, C)$ shows the amount of mutual information between attribute x and the label of class c. $P(x_i)$ and $P(C)$ present marginal probability functions, and $P(x_i, C)$ indicates the joint probability distribution. If two random variables are entirely independent, the mutual information value is 0. The MRMR aims to minimize redundancy (Rd) while maximizing relevance (Re). The top m features related to the class labels are selected using the maximum relevance.

$$\mathrm{Re}(S) = \frac{1}{|S|} \sum_{x_i \in S} \mathrm{MI}(x_i, C) \tag{21}$$

where, $\mathrm{MI}(x_i, C)$ is the mutual information of feature Xi with class C. A Minimum-Redundancy is used to eliminate redundancy between features which are described as follows:

$$\mathrm{Rd}(S) = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} \mathrm{MI}(x_i, x_j) \tag{22}$$

where $\mathrm{MI}(\mathrm{x_i}, \mathrm{x_j})$. is the mutual information of feature x_i. with x_j. A criterion that optimizes relevance and redundancy is called mim redundancy maximum relevance (MRMR). The simplest way to optimize relevance and redundancy to gain an informative subset of features is as follows:

$$\max \emptyset(\mathrm{Re}(S), \mathrm{Rd}(S)) \tag{23}$$

where $\emptyset = (\mathrm{Re}(S)–\mathrm{Rd}(S))$.

4.2 The proposed Binary version of the Horse Herd Optimization Algorithm (BHOA)

This paper introduces a binary version of HOA for feature selection problems called BHOA. Discrete binary search spaces are generally necessary for various applications, such as feature selection. Additionally, problems with continuous values could be turned into binary ones by transforming their variables to [1]. Any binary search space has structure and limitations irrespective of the binary problem types. Each horse in the HOA moves in a continuous search space, with its position vector represented by continuous values. Accordingly, Eq. (2) efficiently implements horse position updates by adding velocity variables to the position vector. In discrete problems, the solutions are limited to binary values like 0 and 1, so the positions cannot be updated using Eq. (2). As a result, a method must be developed to change the horse's location using velocity. The proposed algorithm uses a Transfer Function (TF) and a location update method. The TFs are divided into two groups based on their shapes: S-shaped and V-shaped. These two groups are illustrated in Fig. 2, and Table 1 shows the most commonly used mathematical formulations of the S-shaped and V-shaped TFs [59]. The transfer function produces a probability value according to the velocity of each horse, and this probability value allows continuous positions to convert into binary values.

4.3 A novel X-Shaped Transfer Function

Due to the fact that the existing version of TFs (S-shaped, V-shaped) does not achieve an optimal balance between exploration and exploitation, we utilize a novel X-shaped transfer function for updating the location of each horse [33]. The X-shaped transfer function is used in optimization algorithms to convert continuous search space into a binary one. The intuition behind this function lies in its ability to effectively map continuous values to binary states (0 or 1). It's called 'X-shaped' because of the graphical representation of the function, which resembles an 'X' when plotted. This shape is achieved by combining two S-shaped curves in opposite directions. The function provides a balanced mechanism for updating positions in binary search spaces, ensuring efficient exploration and exploitation. This is particularly useful in problems where binary decision-making is required, as it offers a method to handle continuous data in binary optimization algorithms. As illustrated in Fig. 3, the X-shaped TF includes two functions to enhance exploration and exploitation. Two new positions are generated using Eqs. (24) and (26). The best solution is selected by Eq. (27), then compared to the previous one.

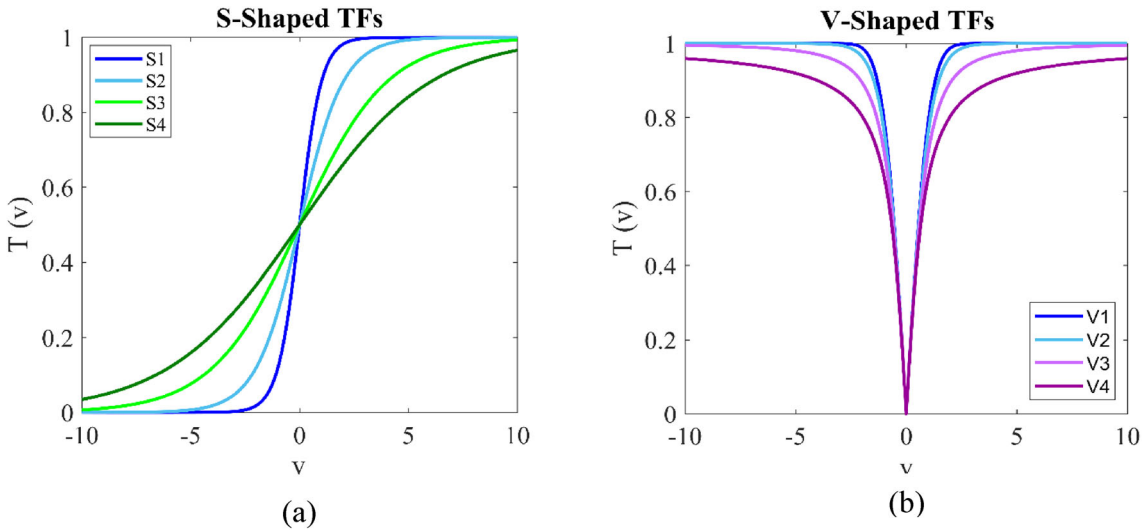

Fig. 2 Transfer functions groups **a** S-Shaped and **b** V-Shaped

Table 1 S-Shaped and V-Shaped transfer functions

S-Shaped		V-Shaped	
Name	Transfer function	name	Transfer function
S_1	$T(v) = \frac{1}{1+e^{-2x}}$	V_1	$T(v) = \left\lvert \text{erf}\left(\frac{\sqrt{\pi}}{2}v\right)\right\rvert$
S_2	$T(v) = \frac{1}{1+e^{-x}}$	V_2	$T(v) = \lvert\tanh(v)\rvert$
S_3	$T(v) = \frac{1}{1+e^{-\left(\frac{x}{2}\right)}}$	V_3	$T(v) = \left\lvert (v)/\sqrt{1+v^2}\right\rvert$
S_4	$T(v) = \frac{1}{1+e^{-\left(\frac{x}{3}\right)}}$	V_4	$T(v) = \left\lvert \frac{2}{\pi}\text{arc}\tan\left(\frac{2}{\pi}v\right)\right\rvert$

$$W_1(v) = \frac{1}{1+e^{-v}} \tag{24}$$

$$D_i(t+1) = \begin{cases} 1\text{if} \quad \text{rand}_1 < W_1(t+1) \\ 0 \text{ if} \quad \text{rand}_1 \geq W_1(t+1) \end{cases} \tag{25}$$

$$W_2(v) = \frac{1}{1+e^{v}} \tag{26}$$

$$G_i(t+1) = \begin{cases} 1 \text{ if } \text{rand}_1 > W_2(t+1) \\ 0 \text{ if } \text{rand}_1 \leq W_2(t+1) \end{cases} \tag{27}$$

$$Z_i(t+1) = \begin{cases} D_i \text{ if } \text{Fitness}(D_i) < \text{Fitness}(G_i) \\ G_i \text{ if } \text{Fitness}\ (D_i) \geq \text{Fitness}(G_i) \end{cases} \tag{28}$$

where, D_i and G_i are the binary value obtained by Eq. (24) and Eq. (26) respectively, and rand_1 and rand_2 are random numbers between 0 and 1. If $Z_i(t+1)$ has a higher fitness value than the current position $(X(i))$, Z_i. will be considered as the new position. Otherwise, we employ a crossover operator on the current location (X(i)) and Z_i to generate two children. The best crossover result determines the next position. The pseudo-code of X-shaped TF is shown in Algorithm 2.

4.4 The proposed hybrid approach based on the BHOA for Gene Selection

This section introduces MRMR-BHOA, a novel hybrid gene selection method that combines BHOA and MRMR to solve feature selection problems. This work aims to find an optimal gene subset regarding the maximum accuracy and the smallest number of genes. Furthermore, the proposed method significantly shortens time complexity and eliminates irrelevant genes. MRMR method is employed to

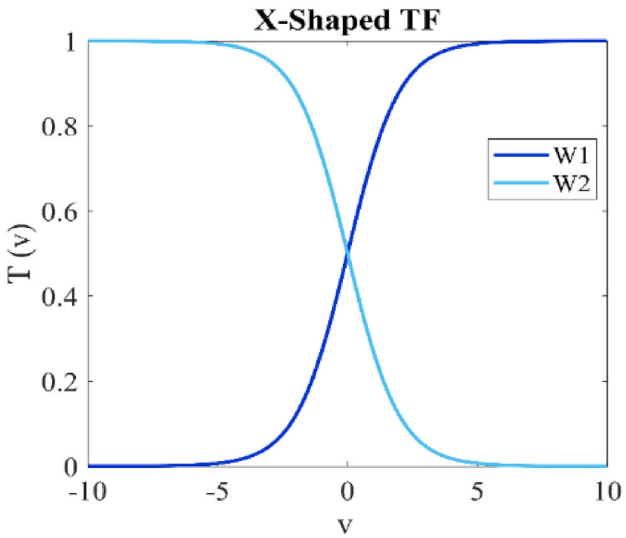

Fig. 3 X-shaped transfer function

minimize redundant genes in the initial stage, followed by BHOA, which is applied to select the most efficient genes. Pseudo-code for the proposed algorithm can be found in algorithm 3. Due to the problem of whether to select a given gene, if the horse position is zero, the corresponding gene is not chosen. If the horse position is one, the corresponding gene is selected. The feature selection mechanism is shown in Fig. 4. The SVM classifier evaluates the selected genes subset using the K-Fold-Cross Validation method. In BHOA, the fitness function assesses each horse's position. The fitness value is calculated according to the objective function below, considering the number of the selected genes and the classification accuracy. The objective function is described as follows:

$$\text{Fitness} = \alpha * \text{ACC} + \beta \frac{|N - S|}{|N|} \tag{29}$$

where, ACC indicates the accuracy of classification (obtained by using the SVM classifier), S and N are the number of selected genes and the total number of genes, respectively. In addition, the two parameters, α, and β, correspond to the importance of classification quality and subset length, respectively. α is in the range [0, 1] and $\beta = (1 - \alpha)$. A framework of the proposed approach to the gene selection problem is illustrated in Fig. 5.

Algorithm 2: X-Shaped transfer function

for i = 1: total number of horses **do**
  Computing the velocity (V_i) of each horse by Equation (2)
  Calculate $W_1(V_i) = \frac{1}{1 + e^{-v_i}}$
  if $rand_1 < W_1(V_i)$ **then**
    $D_i = 1$
  else
    $D_i = 0$
  end if
  Calculate $W_2(V_i) = \frac{1}{1 + e^{v_i}}$
  if $rand_2 < W_2(V_i)$ **then**
    $G_i = 1$
  else
    $G_i = 0$
  end if
  Calculate the new position X(t+1):
  if $f(D_i) < f(G_i)$ **then**
    $Z(t+1) = D_i$
  else
    $Z(t+1) = G_i$
  end if
  if $f(Z(t+1)) > f(X(t))$ **then**
    $X(t+1) = Z(t+1)$
  else
    $[child_1, child_2] = \text{Crossover}(Z(t+1), X(t))$
    $X(t+1) = \text{best}(child_1, child_2)$
  end if
end for

Algorithm 3: MRMR-BHOA

Define input parameters
Generate random binary positions for *n* horses
Evaluate fitness of each horse's location
Generate Global Matrix based on the horses' location and their fitness value
while *(the stopping criterion is not satisfied)* **do**
    for i = 1: total number of horses **do**
        Sort the locations of horses in ascending order depending upon their fitness value
        Calculate the mean position by Equation (9)
        Calculate the good position by Equation (12)
        Calculate the bad position by Equation (15)
        Determining alpha, beta, gamma, and delta horses
        Calculate $W_1(V_i) = \frac{1}{1+e^{-v_i}}$
        if $rand_1 < W_1(V_i)$ **then**
            $D_i = 1$
        else
            $D_i = 0$
        end if
        Calculate $W_2(V_i) = \frac{1}{1+e^{v_i}}$
        if $rand_2 < W_2(V_i)$ **then**
            $G_i = 1$
        else
            $G_i = 0$
        end if
        Calculate the new position X(t+1):
        if $f(D_i) < f(G_i)$ **then**
            $Z(t+1) = D_i$
        else
            $Z(t+1) = G_i$
        end if
        if $f(Z(t+1)) > f(X(t))$ **then**
            $X(t+1) = Z(t+1)$
        else
            $[child_1, child_2] = Crossover(Z(t+1), X(t))$
            $X(t+1) = best(child_1, child_2)$
        end if
        if new fitness value < old fitness value **then**
            set new position as the best position
            set new fitness value as the best fitness value
        end if
    end for
end while
return the best position
return the best fitness value

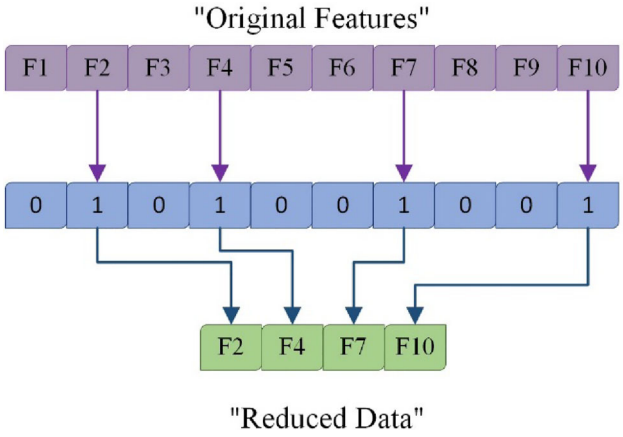

Fig. 4 The mechanism of selecting features

5 Experimental result and discussion

5.1 Experimental setup

This section examines the proposed method on ten microarray datasets, consisting of DLBCL, Leukemia, SRBCT, Ovarian, Colon, Prostate, Lung, Brain-1, MLL, and Lymphoma that belong to the biomedical domain. Table 2 elaborates on the details of each dataset concerning the number of genes and objects. The ten-fold cross-validation method, which randomly divides each dataset into two subsets called training and testing, was utilized to test the efficiency of the results. All approaches were repeated 20 times to achieve statistically significant results. Furthermore, implementing the proposed approaches was programmed using MATLAB 2021b, Intel Core i7, and 12 GB of RAM.

5.2 Parameters tuning

The proposed MRMR-BHOA approach is compared to various widely used algorithms, such as Gray Wolf Optimization (GWO), Whale Optimization Algorithm (WOA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Firefly algorithm. Listed below are the parameters set for competing algorithms in Table 4. To ensure fair comparisons, we assign a population size of 35 and a maximum number of iterations of 60, respectively. In addition, we chose α and β in accordance with the literature (i.e., 0.99 and 0.01, respectively) [60]. The total features in a dataset determine the size of its search space. Furthermore, the efficiency of the results is also proven by statistical analysis.

5.3 Measurement criteria

The classification accuracy is the first evaluation criterion used in this work. Generally speaking, the accuracy of classification measures indicates whether a classifier correctly identifies each feature label. The results of these evaluations are computed based on a confusion matrix given in Table 3. Accuracy can be calculated by comparing rightly classified samples to the entire sample pool, as described below:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} \tag{30}$$

where TP and TN represent true positives and true negatives, and false positives and false negatives are shown *by FP and FN,* respectively. As a result, we examine the capability of the proposed method using the SVM classifier, which is used as an accuracy value in the fitness function. Matthews Correlation Coefficient (MCC), F-measure, and Area under ROC Curve (AUC) are also employed to assess the effectiveness of the proposed approach. The following Equations describe these criteria:

$$\text{MCC} = \frac{\text{TN} \times \text{TP} - \text{FN} \times \text{FP}}{\sqrt{(\text{TP} + \text{FP}) + (\text{TP} + \text{FN}) + (\text{TN} + \text{FP}) + (\text{TN} + \text{FN})}} \tag{31}$$

$$\text{F - Measure} = 2^{*} \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{32}$$

5.4 Experimental results and analysis

In this paper, we conduct a comparative analysis of the Binary Horse Herd Optimization Algorithm (BHOA) against contemporary methodologies, focusing on evaluating the impact of the binary HOA and the X-shaped transfer functions in the context of feature selection. Our study is structured into three critical phases: Initially, we assess the performance of four distinct classifiers across various datasets to identify the most suitable one for evaluating our approach. Subsequently, we apply and contrast different filter methods to ascertain the most efficient technique for dimensionality reduction. The last step is to compare the X-shaped TF to the S-shaped and V-shaped TF and demonstrate its significance. Four diverse classifiers (i.e., KNN, SVM, DT, and NB) were tested on ten biological datasets to determine which classification algorithm is more effective than the others. The comparison of various classifiers can be found in Table 5, and Figs. 6 and 7 illustrate the performance of different classifiers on ten datasets regarding accuracy and F-measure. As a result of their close results, we decided to use SVM and KNN classifiers as evaluators during the filter method step.

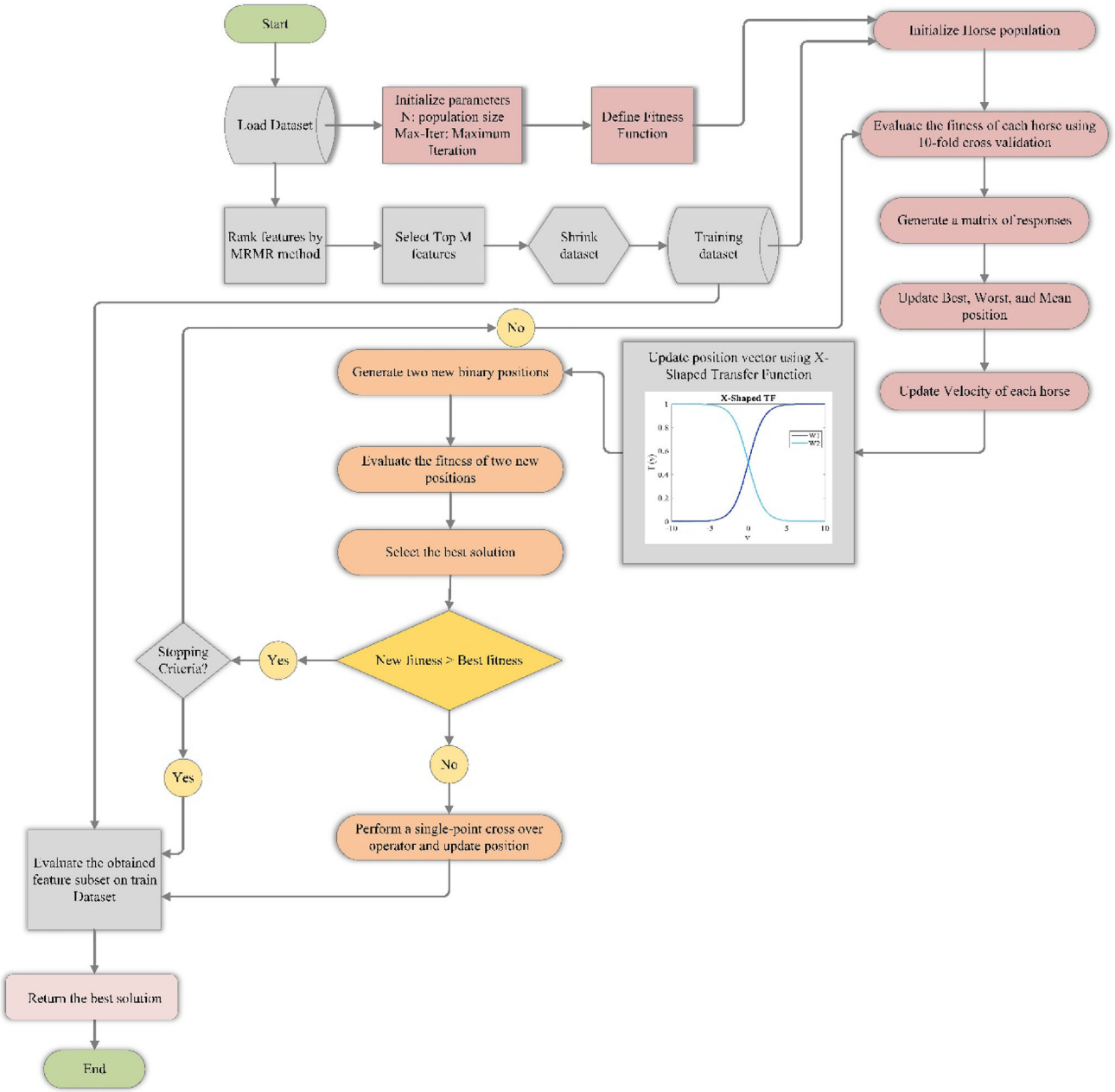

Fig. 5 A framework of the proposed approach for the gene selection problem

5.4.1 Assessment of the performance of various filter techniques

In the following step, we compared the results of four different filter techniques to determine the most effective filter method for identifying informative genes to use as input to the proposed wrapper approach. As a result, we used the Relief, Chi-square, Laplacian, and MRMR methods to select the best 50 and 100 genes, which were then evaluated using the SVM and KNN classifiers. Table 6 presents a comparison of different filters. The results in Table 6 clearly show that the MRMR method outperforms other filter methods in terms of accuracy on all microarray datasets. Laplacian has the worst performance across all datasets. MRMR technique achieves higher classification accuracy on 7 out of 10 datasets for the top 50 genes. On the MLL, Leukemia, and Ovarian datasets, the best performance is obtained for the top 100 genes by the MRMR method.

It should be noted that the SVM classifier with linear kernel achieves the highest efficiency on 8 out of 10 datasets. In contrast, the KNN classifier obtains better results on the remaining datasets (i.e., Colon, Prostate). Figures 8 and 9

Table 2 Dataset description

Datasets	No. of genes	No. of instances	No. of classes
Lymphoma	4026	66	3
Prostate	10,509	102	2
Brain-1	5920	90	5
DLBCL	4026	47	2
Colon	2000	62	2
Leukemia	7129	72	2
SRBCT	2308	83	4
Lung	12,600	203	5
Ovarian	15,154	253	2
MLL	12,582	72	3

Table 3 The parameters settings

Algorithm	Parameter	Explanation	Value
BHOA	W	Reduction factor	0.9
	pN	Percent of best horses	0.1
	qN	Percent of worst horses	0.2
PSO	C_1, C_2	Acceleration coefficient	1.5, 2
	W	Inertia weight	0.9
GA	P_s	Selection mechanism	Roulette wheel
	P_c	Crossover ratio	0.7
	P_m	Mutation ratio	0.1
Firefly	α	Attractiveness coefficient	0.2
	β	Absorption coefficient	1
	γ	Light absorption coefficient	1
GWO	a	Coefficient	[2, 0]
ACO	α	Pheromone	1
	β	Visibility	2
	ρ	Evaporation of pheromone	0.5
WOA	a	Coefficient	[2, 0]
	b	Logarithmic spiral constant	1

compare the performance of two different classifiers on the top 50 and 100 genes of each dataset using the MRMR filter method.

Assessment of the performance of BHOA using S-shaped and V-shaped TFs The effectiveness of binary HOA (BHOA) on feature selection is investigated using eight different transfer functions (S-shaped and V-shaped, as mentioned in Table 1) and an SVM classifier with the linear kernel as an evaluator. Table 7 presents the results of BHOA based on S-shaped and V-shaped TFs in terms of best, mean, worst, and standard deviation of classification accuracy. Also, Table 8 reports the minimum number of selected genes using each transfer function. As shown in Tables 7 and 8, S-shaped groups achieved the same results in 5 cases, and S1 showed good performance in 3 out of the rest of the datasets. Moreover, V1 obtained the highest classification accuracy in 5 out of 10 datasets in the V-shaped family. Overall, the best results of S-shaped families were reported by S1, and the V1 TF achieved higher average accuracy and a lower number of genes than other V-shaped TFs on all datasets.

By comparing the results of S1 and V1 in terms of average accuracy, it can be found that they obtained the same results for five datasets, and V1 achieved higher performance on 3 out of 5 remaining datasets. Table 8 shows that S1 and V1 obtained nearly the same results for selected genes. However, the results of S1 are better than V1 on 8 out of 10 datasets. Since the primary purpose of gene selection is to decrease the number of selected genes while maximizing classification accuracy, it can be concluded that BHOA using S1 performed well because they achieved the highest classification accuracy while using fewer genes. Additionally, comparing the standard deviations between S1 and V1 demonstrates that the S1 transfer function provided smoother results with enhanced stability.

Table 4 Confusion matrix

		Actual condition	
		Actual positive	Actual negative
Output of classifier	Classify positive	TP	FP
	Classify negative	FN	TN

Assessment of the impact of X-shaped TF on BHOA To investigate the impact of the novel X-Shaped TF on the performance of the BHOA, we should compare the obtained results by BHOA and eight TFs and results by BHOA and X-Shaped TF. The best, mean, worst, and standard deviation (STD) of classification accuracy are common criteria used to evaluate the effectiveness of the proposed approach, and an average of selected genes is measured and displayed in Table 9. Then, an SVM classifier with ten-fold cross-validation is employed to evaluate the selected genes. It can be seen from Table 9 that the BHOA using X-shaped TF is

Table 5 The performance of four different classifiers on ten biological datasets

Dataset	Criteria	Classifier			
		SVM	KNN	NB	DT
Lymphoma	Accuracy	100	96.9	92.4	81.8
	F-Measure	100	97.0	92.0	81.8
	MCC	100	93.2	83.0	65.7
	AUC	100	95.4	87.5	89.8
Prostate	Accuracy	91.1	84.3	62.7	87.2
	F-Measure	91.2	84.3	59.9	87.3
	MCC	82.4	69.2	28.8	74.7
	AUC	91.2	84.4	62.7	97.6
Brain-1	Accuracy	80.6	82.9	86.8	81.1
	F-Measure	79.4	78.9	86.9	79.9
	MCC	91.7	90.3	90.5	91.1
	AUC	79.4	78.8	86.8	79.8
DLBCL	Accuracy	91.4	76.5	95.7	82.9
	F-Measure	91.5	55.4	95.7	82.9
	MCC	83.3	76.0	91.5	66.1
	AUC	91.6	76.3	98.8	76.6
Colon	Accuracy	80.6	75.8	58.0	80.6
	F-Measure	80.4	75.7	58.9	80.4
	MCC	57.0	46.6	19.9	57.0
	AUC	77.8	73.1	64.8	66.5
Leukemia	Accuracy	97.2	87.5	100	73.6
	F-Measure	97.2	87.3	100	74.1
	MCC	93.9	72.0	100	44.9
	AUC	96.9	84.8	100	69.4
SRBCT	Accuracy	97.6	84.33	98.8	80.7
	F-Measure	97.6	83.6	98.8	80.6
	MCC	96.7	78.9	98.2	73.2
	AUC	98.1	96.9	99.1	87.5
Lung	Accuracy	93.6	87.7	80.8	85.2
	F-Measure	93.4	85.3	81.4	84.7
	MCC	86.1	94.8	65.5	71.1
	AUC	90.9	94.7	85.3	88.4
Ovarian	Accuracy	100	96.0	88.1	95.6
	F-Measure	100	96.0	88.2	95.6
	MCC	100	91.4	74.5	90.5
	AUC	100	95.5	93.0	95.7
MLL	Accuracy	95.8	83.3	95.8	86.1
	F-Measure	95.8	83.8	95.9	86.0
	MCC	93.8	76.7	93.7	78.8
	AUC	96.9	88.1	96.8	87.4

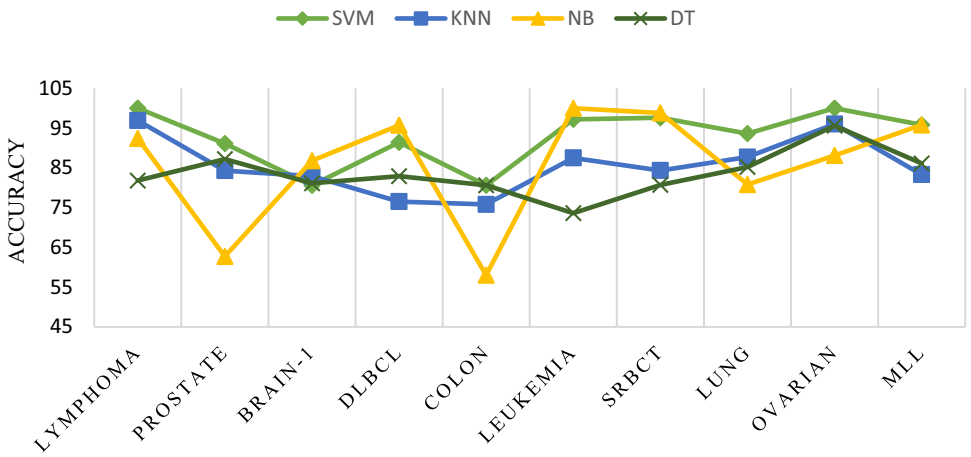

Fig. 6 The performance of four classifiers on ten datasets in terms of accuracy

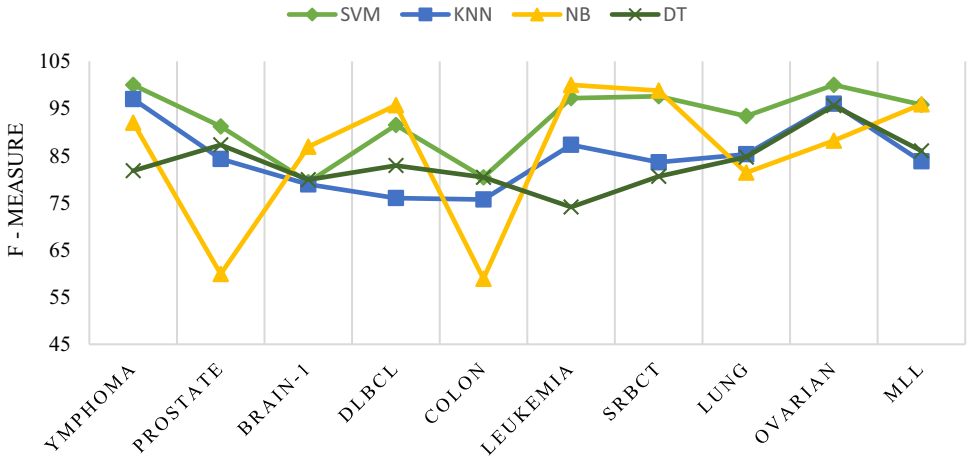

Fig. 7 The performance of four classifiers on ten datasets in terms of F-measure

able to obtain 100% average accuracy for six datasets, which are Lymphoma, DLBCL, SRBCT, Leukemia, Ovarian, and MLL. Also, it has achieved 98.48%, 98.66%, 97.72%, and 96.45% classification accuracy for the Colon, Lung, Prostate, and Brain-1 datasets, respectively. Furthermore, the proposed approach showed superior performance in terms of selected genes. In other words, the proposed method has the capability to gain higher accuracy with a minimum number of selected genes. Table 10 shows the best results of X-Shaped, S1, and V1 TFs to demonstrate the strength of X-Shaped TF. According to Table 10, it can be detected that BHOA using X-shaped TF shows superiority in comparison to S1 and V1 in terms of classification accuracy and the number of selected genes. It can be concluded that X-Shaped TF outperforms traditional transfer functions. Table 11 summarizes the best subset of genes for each dataset. Figures 10 and 11 depict the convergence behavior of the proposed method when different transfer functions are used.

5.5 Comparative results analysis between the proposed method and other meta-heuristic algorithms

This section provides a comparison between the BHOA algorithm and several popular meta-heuristic algorithms, including Gray Wolf Optimization (GWO), Whale Optimization Algorithm (WOA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Genetic Algorithm (GA), and Firefly algorithm, to demonstrate the efficacy of the proposed approach using X-Shaped transfer function. All experiments were carried out 20 times independently to acquire confident results. Table 4 displays the parameters that were used to run these algorithms. Furthermore, the obtained results are outlined in Table 12 concerning the average classification accuracy, standard deviation, and the number of selected genes. Table 12 shows that the BHOA and WOA have achieved 100% accuracy on 6 out of 10 datasets. However, BHOA outperforms WOA for all datasets regarding the minimum number of selected genes. Although BHOA provided similar results to PSO, GA, GWO, and WOA in terms of accuracy in the Ovarian dataset, it could identify

Table 6 Average classification accuracy (%) of top 50 and 100 genes obtained by different filter methods

Dataset	Classifier	50 genes				100 genes			
		MRMR	Relieff	Chi-square	Laplacian	MRMR	Relieff	Chi-square	Laplacian
Lymphoma	SVM	100	100	98.48	96.90	100	100	98.48	98.33
	KNN	98.33	100	96.96	95.00	98.33	96.96	92.42	94.22
Prostate	SVM	92.24	92.16	90.19	65.57	91.84	95.09	88.23	68.46
Brain-1	KNN	94.01	94.11	93.14	63.59	85.27	93.13	90.19	62.69
	SVM	84.44	83.26	84.73	68.18	90.52	88.72	90.12	75.73
	KNN	87.77	86.49	87.39	69.44	90.11	87.49	90.06	74.85
DLBCL	SVM	100	97.50	100	69.83	100	100	100	65.84
	KNN	95.74	100	95.74	61.33	100	100	95.74	60.83
Colon	SVM	83.87	83.87	83.91	73.90	82.14	85.36	79.23	72.62
	KNN	85.49	77.41	80.64	58.93	80.92	87.09	79.03	51.42
Leukemia	SVM	95.83	97.51	97.22	64.34	100	98.64	95.83	63.69
	KNN	93.05	96.07	93.25	73.15	98.75	98.61	97.22	73.15
SRBCT	SVM	100	97.63	98.79	81.12	100	100	100	89.16
	KNN	100	100	98.76	70.49	100	100	100	65.67
Lung	SVM	95.57	89.65	88.17	76.88	94.45	89.65	92.61	78.28
	KNN	95.56	92.61	93.10	68.47	94.16	93.59	93.10	68.05
Ovarian	SVM	99.61	98.83	98.81	90.87	99.62	98.02	100	90.41
	KNN	99.61	99.21	99.20	73.96	99.6	98.81	99.60	82.36
MLL	SVM	98.57	94.40	91.66	73.81	100	93.05	94.44	69.21
	KNN	98.46	98.46	95.83	58.51	100	95.83	93.05	58.75

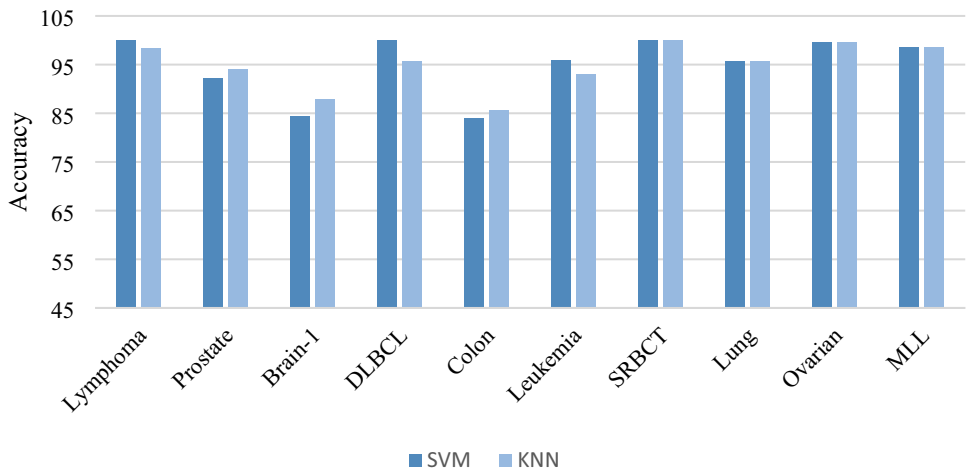

Fig. 8 Comparison between the performance of two different classifiers on the obtained top 50 genes of each dataset by the MRMR filter method

fewer optimal genes when higher accuracy was considered. The same competition can be seen on the SRBCT dataset. BHOA obtained a minimal number of genes compared to PSO, GWO, and WOA, while they obtained 100% accuracy. For the rest of the datasets, BHOA performs better than others regarding average accuracy and selected genes.

Moreover, the standard deviation is low for accuracy and selected genes across all datasets, and it is smoother than other approaches. In summary, the proposed method can more accurately identify the minimal optimal genes. Based on the results of Table 12, it is evident that the proposed approach outperforms the other meta-heuristics used in this study. Figures 12 and 13 compare the proposed method and other meta-heuristic algorithms regarding classification accuracy and the number of selected genes.

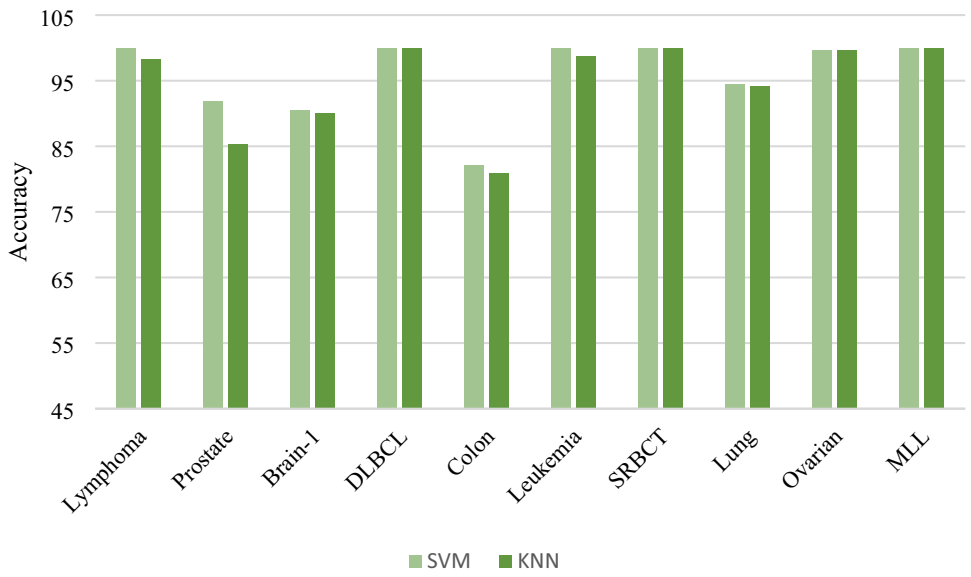

Fig. 9 Comparison between the performance of two different classifiers on the obtained top 100 genes of each dataset by the MRMR filter method

5.6 Evaluation of proposed method with past literature

Due to BHOA's superior performance compared to other meta-heuristic algorithms, experimental studies are expanded to compare MRMR-BHOA's performance with existing state-of-the-art algorithms. Table 13 summarizes the mean accuracy and the number of selected genes (in brackets) achieved by the proposed method and other methods in the literature. The bold type indicates the best results; a "*" means no result was found. As can be seen from Table 13, MRMR-BHOA performed better than all other comparative methods for ten datasets concerning classification accuracy and the minimum number of selected genes.

To elaborate, MRMR-BHOA obtained 100% average accuracy with only two genes in the Lymphoma dataset, while three approaches achieved 100% accuracy with more genes. For the Prostate dataset, BFO [61] and DRF0-CGS [62], as well as the proposed approach, achieved approximately 97.5% accuracy. However, the proposed method identified only 5.4 genes, so it outperformed other techniques. In the Brain-1 dataset, DEFS [63] achieved 96.30% accuracy with 30 genes, while the proposed method obtained 96.45% with fewer genes (8.81) than DEFS. The proposed algorithm selected only 2.8 genes with 100% accuracy on the DLBCL dataset. The proposed algorithm provided higher accuracy for the colon dataset with fewer genes (7.36). All techniques have achieved 100% accuracy in the Leukemia dataset. On the other hand, the proposed method selected only 2.6 genes. Although 3 out of 5 methods obtained 100% accuracy, similar to MRMR-BHOA, the proposed method chose fewer genes (5.53) than others on the SRBCT dataset. Moreover, DRF0-CGS and MRMR-BHOA performed better than the other methods in the lung dataset; despite this, MRMR-BHOA identified few genes (8.36). In the two remaining datasets, Ovarian and MLL, three cases of five methods have attained 100% accuracy, but the MRMR-BHOA algorithm has selected the fewest genes. For the Ovarian and MLL datasets, the proposed method reached 100% accuracy with only 3 and 4 generations, respectively. Overall, the findings indicate that the proposed methodology can produce fewer gene sets with more accurate classification than other approaches on the whole dataset.

5.7 Statistical results

To substantiate the existence of significant differences between the results from multiple approaches, we used the Friedman test with 5% significance, a commonly used non-parametric statistical method [79]. Table 14 generally ranks all methods separately on each dataset, with the best algorithm receiving a rank of 1 and the worst algorithm receiving the last place. Tests are performed on accuracy and selected genes from each dataset. Tables 14 and 15 present the average rank of algorithms across all datasets using Friedman's statistics regarding classification accuracy and the number of selected genes. According to the results, the proposed method ranks first in accuracy and selected genes. Based on the reported results, Friedman's statistics are 30.89999 and 40.5, and the p-values calculated in this experiment are 0.0000026488 and 0.0000000363 for accuracy and selected genes, respectively. In this case, the null hypothesis is rejected, indicating a difference between the performances of the compared algorithms. Since we found a statistically notable outcome, we used a post-hoc test for

Table 7 A comparison between the S-shaped and V-shaped transfer functions regarding the best, the mean, the worst, and STD of classification accuracy (%)

Dataset	Performance	S-Shaped TF				V-Shaped TF			
		S1	S2	S3	S4	V1	V2	V3	V4
Lymphoma	Best	100	100	100	100	100	100	100	100
	Mean	100	100	100	100	100	100	100	100
	Worst	100	100	100	100	100	100	100	100
	STD	0	0	0	0	0	0	0	0
Prostate	Best	95.61	95.5	95.16	94.35	93.59	92.98	93.71	92.99
	Mean	95.59	95.02	95.03	94.01	93.12	92.61	92.82	92.05
	Worst	95.28	94.51	94.78	93.49	92.46	92.01	92.40	91.69
	STD	1.12	1.03	0.76	1.67	1.72	0.54	1.36	0.89
Brain-1	Best	93.77	93.87	94.77	94.55	94.77	93.77	94.66	93.77
	Mean	92.92	93.03	93.65	93.39	94.44	93.26	93.86	93.37
	Worst	92.55	92.31	92.44	91.55	93.64	92.66	93.44	92.44
	STD	0.48	0.73	0.82	1.31	0.45	0.55	0.47	0.53
DLBCL	Best	100	100	100	100	100	100	100	100
	Mean	100	100	100	100	100	100	100	100
	Worst	100	100	100	100	100	100	100	100
	STD	0	0	0	0	0	0	0	0
Colon	Best	98.33	96.91	96.89	97.33	96.91	97.14	97.51	98.33
	Mean	96.99	96.67	96.47	96.75	96.85	96.82	97.12	97.28
	Worst	95.47	96.47	95.23	95.47	96.61	96.62	96.93	96.9
	STD	1.01	0.15	0.71	0.73	0.14	0.19	0.27	0.61
Leukemia	Best	100	100	100	100	100	100	100	100
	Mean	100	99.26	99.10	99.02	99.27	99.26	99.17	99.02
	Worst	100	98.75	98.74	98.64	98.75	98.45	98.36	98.12
	STD	0	0.64	0.60	0.71	0.64	0.69	0.71	0.73
SRBCT	Best	100	100	100	100	100	100	100	100
	Mean	100	100	100	100	100	100	100	100
	Worst	100	100	100	100	100	100	100	100
	STD	0	0	0	0	0	0	0	0
Lung	Best	97.54	98.57	97.57	98.07	97.59	97.59	98.07	97.52
	Mean	97.08	98.15	97.46	97.45	97.47	97.35	97.56	97.35
	Worst	96.63	97.57	97.11	97.04	97.11	97.07	97.04	97.12
	STD	0.32	0.41	0.19	0.42	0.25	0.25	0.36	0.37
Ovarian	Best	100	100	100	100	100	100	100	100
	Mean	100	100	100	100	100	100	100	100
	Worst	100	100	100	100	100	100	100	100
	STD	0	0	0	0	0	0	0	0
MLL	Best	100	100	100	100	100	100	100	100
	Mean	100	100	100	100	100	100	100	100
	Worst	100	100	100	100	100	100	100	100
	STD	0	0	0	0	0	0	0	0

Table 8 A comparison between the S-Shaped and V-Shaped transfer functions regarding the average and STD of the selected genes

Dataset	Measure	S-Shaped TF				V-Shaped TF			
		S1	S2	S3	S4	V1	V2	V3	V4
Lymphoma	Avg.	17	17	17.16	18.8	15.2	17.33	15.6	17.12
	STD	1.33	1.62	2.41	1.30	2.2	1.2	2.1	2.5
Prostate	Avg.	23.6	23.2	24.6	23.79	24.89	25.07	25.73	25.40
	STD	1.94	2.16	1.98	1.49	2.06	2.16	1.65	1.01
Brain-1	Avg.	26	24.4	22.8	23.8	25.8	22.4	21.6	22.8
	STD	1.22	1.14	1.64	0.84	1.31	1.52	2.07	1.92
DLBCL	Avg.	11.5	17.33	19.57	19.14	18.51	17.42	17	18.43
	STD	1.35	1.86	0.97	1.06	1.1	1.12	2.7	1.9
Colon	Avg.	22.4	24.6	22.2	24.2	27.4	26.83	24.2	24.8
	STD	1.89	1.52	0.84	1.48	1.67	1.92	1.09	1.78
Leukemia	Avg.	15.7	21.7	21.28	22.38	18.16	20.44	20.46	21.43
	STD	1.56	2.08	2.14	2.56	1.9	0.9	0.9	0.8
SRBCT	Avg.	11.8	17.75	19.2	19.33	17.33	18.42	18.87	18.68
	STD	0.87	1.28	1.31	1.21	1.6	1.5	1.4	1.3
Lung	Avg.	23	25.2	26	25	27	23	23.8	24.6
	STD	1.41	0.83	1.41	1.22	1.58	2.91	1.92	1.36
Ovarian	Avg.	17.66	17.87	17.94	17.89	18.77	18.81	18.75	18.82
	STD	2.06	1.64	1.87	1.59	1.3	1.3	1.5	1.4
MLL	Avg.	19.11	19.44	20.6	19.12	18.71	18.8	18.84	19.4
	STD	1.05	1.66	0.89	1.9	1.9	1.8	1.7	1.82

Table 9 Experiment results of MRMR-BHOA in terms of accuracy, the number of selected genes, and running time

Dataset	Accuracy (%)				Number of genes		Running time (s)
	Best	Mean	Worst	STD	Mean	STD	
Lymphoma	100	100	100	0	2	0	44.28
Prostate	98.45	97.72	97.12	0.45	5.4	1.26	112.1
Brain-1	96.89	96.45	95.75	0.49	8.81	1.25	94.34
DLBCL	100	100	100	0	2.8	0.63	31.85
Colon	100	98.48	96.66	0.87	7.36	1.61	53.91
Leukemia	100	100	100	0	2.6	0.51	34.88
SRBCT	100	100	100	0	5.53	1.05	60.96
Lung	99.21	98.66	98.49	0.23	8.36	2.20	114.82
Ovarian	100	100	100	0	3	0	52.53
MLL	100	100	100	0	4.1	0.85	48.13

Table 10 A comparison between the proposed method with S-Shaped and V-Shaped transfer functions regarding the average accuracy (%) and the number of selected genes

Dataset	Performance	Transfer function		
		S-Shaped	V-Shaped	X-Shaped
Lymphoma	Accuracy	100 ± 0	100 ± 0	**100 ± 0**
	#Gene	17 ± 1.33	15.2 ± 2.2	**2 ± 0**
Prostate	Accuracy	95.59 ± 1.12	93.12 ± 1.72	**97.72 ± 0.45**
	#Gene	23.6 ± 1.94	24.89 ± 2.06	**5.4 ± 1.26**
Brain-1	Accuracy	92.92 ± 0.48	93.26 ± 0.55	**96.45 ± 0.49**
	#Gene	26 ± 1.22	25.8 ± 1.31	**8.81 ± 1.25**
DLBCL	Accuracy	100 ± 0	100 ± 0	**100 ± 0**
	#Gene	11.5 ± 1.35	18.51 ± 1.1	**2.8 ± 0.63**
Colon	Accuracy	96.99 ± 1.01	96.82 ± 0.14	**98.48 ± 0.87**
	#Gene	22.4 ± 1.89	27.4 ± 1.67	**7.36 ± 1.61**
Leukemia	Accuracy	100 ± 0	99.17 ± 0.64	**100 ± 0**
	#Gene	15.7 ± 1.56	18.16 ± 1.9	**2.6 ± 0.51**
SRBCT	Accuracy	100 ± 0	100 ± 0	**100 ± 0**
	#Gene	11.8 ± 0.87	17.33 ± 1.6	**5.53 ± 1.05**
Lung	Accuracy	97.08 ± 0.32	97.47 ± 0.25	**98.66 ± 0.23**
	#Gene	23 ± 1.41	27 ± 1.58	**8.36 ± 2.20**
Ovarian	Accuracy	100 ± 0	100 ± 0	**100 ± 0**
	#Gene	17.66 ± 2.06	18.77 ± 1.3	**3 ± 0**
MLL	Accuracy	100 ± 0	100 ± 0	**100 ± 0**
	#Gene	19.11 ± 1.05	18.71 ± 1.9	**4.1 ± 0.85**

The bold values signify the best accuracy and number of genes achieved by the specific transfer function for each dataset

Table 11 The obtained best subset of genes by the proposed approach for each dataset

Dataset	#Genes	Gene index	Gene name
Lymphoma	2	8, 47	V2750, V3754
Prostate	6	1, 3, 10, 11, 24, 45	–
Brain-1	6	1, 10, 15, 37, 41, 50	V2532, V5066, V3586, V2331, V5199, V1955
DLBCL	3	6, 12, 47	V1281, V1291, V3020
Colon	8	2, 5, 12, 21, 24, 35, 42, 45	–
Leukemia	3	1, 19, 27	V1834, V1685, V1630
SRBCT	5	10, 12, 14, 17, 18	V1536, V255, V2050, V335, V417
Lung	7	8, 13, 21, 22, 29, 46, 47	V8457, V8531, V8125, V4525, V9164, V12375, V5486
Ovarian	3	1, 17, 50	V1680, V182, V2241
MLL	4	3, 9, 13, 17	–

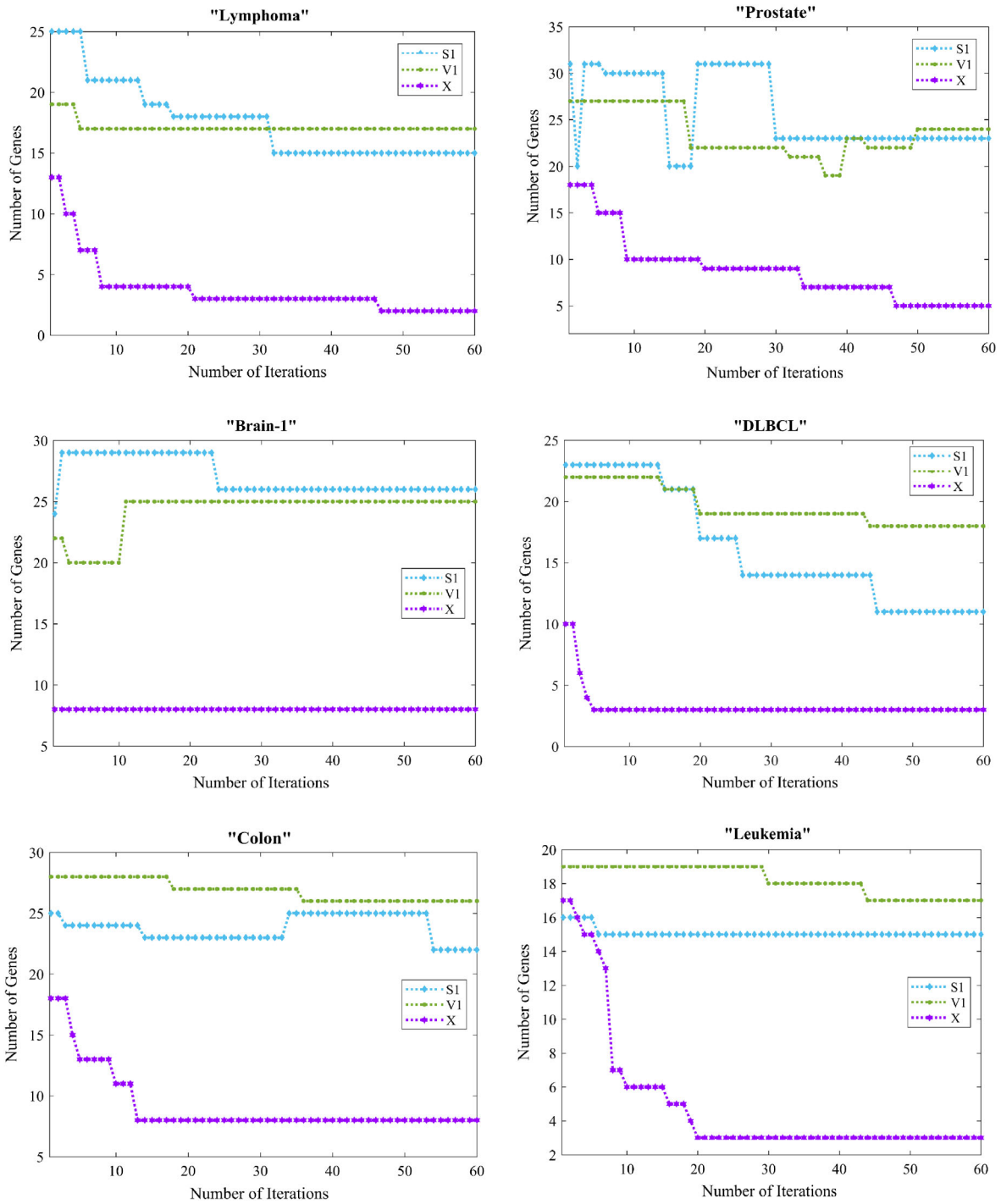

Fig. 10 The convergence behavior of the proposed method using X-Shaped, S-Shaped, and V-Shaped TFs

pairwise comparisons to find the precise cause of our differences. The p-value results of the post-hoc test between the classification accuracy and the minimum selected genes acquired by the BHOA and those of the PSO, GA, Firefly, ACO, GWO, and WOA meta-heuristic algorithms are exhibited in Table 16 and 17. A *p*-value less than 0.05 confirms that the BHOA considerably outperforms its corresponding optimization algorithm. As a result, the proposed approach is statistically superior to most existing methods. A comparison between the ranking of existing algorithms and the proposed approach by the Friedman test is illustrated in Fig. 14.

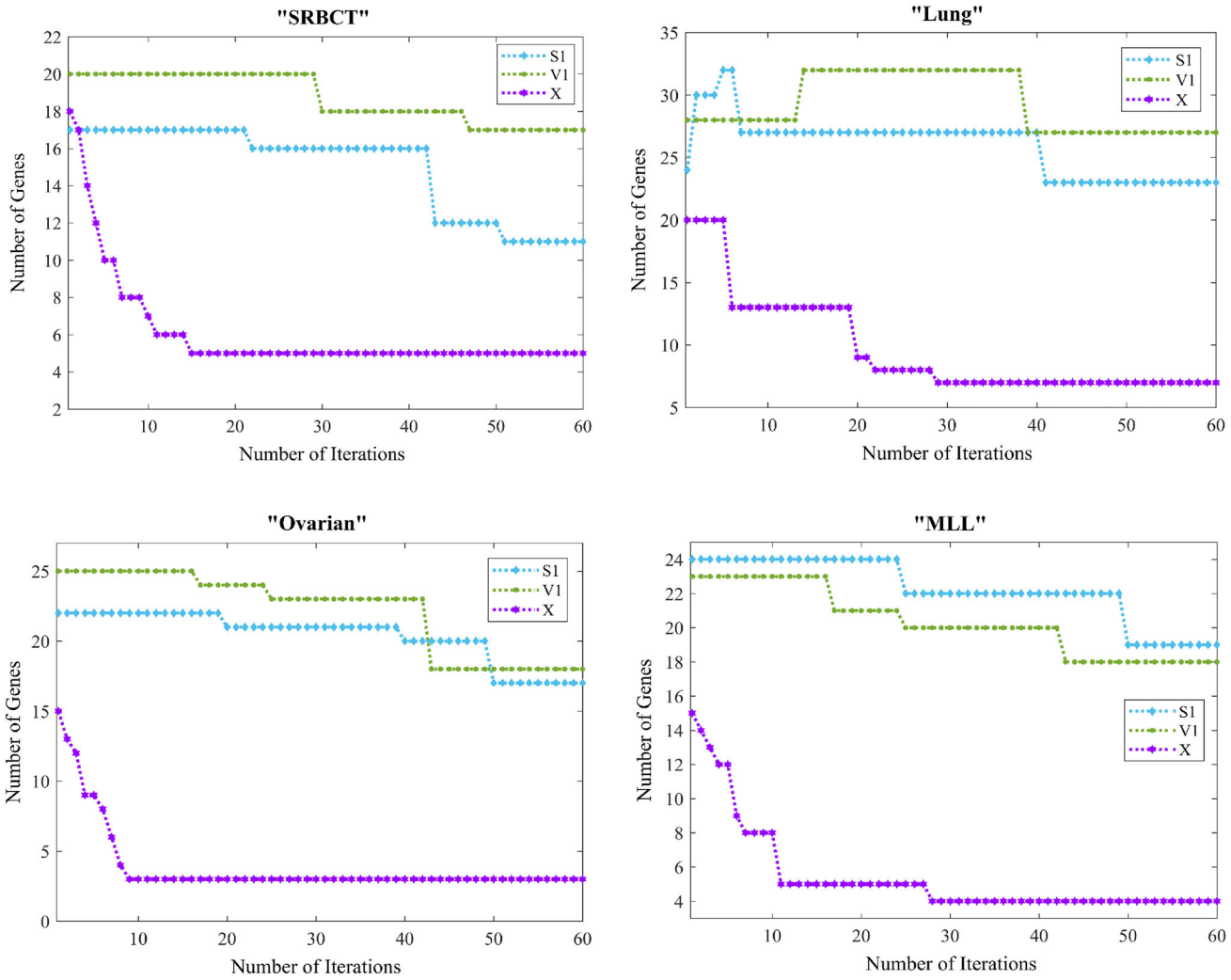

Fig. 11 The convergence behavior of the proposed method using X-Shaped, S-Shaped, and V-Shaped TF

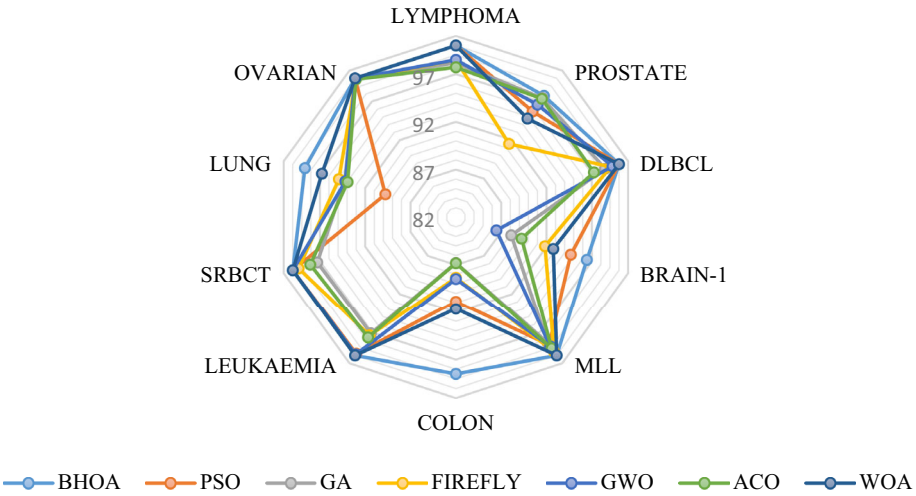

Fig. 12 Comparison of proposed algorithm versus other meta-heuristic algorithms regarding the accuracy

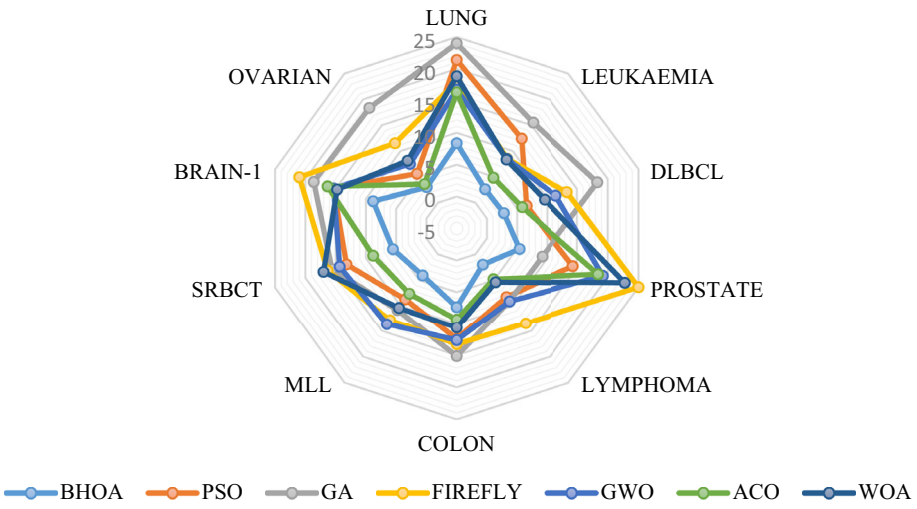

Fig. 13 Comparison of proposed algorithm versus other meta-heuristic algorithms regarding selected genes

6 Conclusion

Over recent decades, biologists and life scientists have been deeply engaged in the challenging task of identifying critical genes related to cancer within microarray datasets. This challenge stems from the excessive number of genes and limited sample sizes in these datasets. Our study addresses gene selection as a binary problem, introducing the Binary Horse Herd Optimization Algorithm (BHOA), a binary adaptation of a novel swarm intelligence algorithm inspired by the behavior of horses. This paper proposes a hybrid method combining BHOA with the Minimum Redundancy Maximum Relevance (MRMR) approach, leveraging wrapper- and filter-based gene selection techniques. The MRMR method initially filters out redundant and irrelevant genes, focusing on significant gene subsets. Subsequently, we map the continuous search space into binary values using eight transfer functions from the S-shaped and V-shaped families.

A distinctive aspect of our approach is the application of an X-shaped Transfer Function (TF) within the MRMR-BHOA framework, enhancing BHOA's performance in gene selection. The X-shaped TF, employing a crossover operation, aids the optimization algorithm in pinpointing the region likely to contain the optimal solution. In our evaluations, the X-shaped TF outperformed other TFs, achieving higher accuracy and selecting the fewest genes. We assessed the proposed framework using ten microarray datasets, analyzed with a Support Vector Machine (SVM) classifier employing ten-fold Cross-Validation. Comparisons with six popular meta-heuristic algorithms demonstrated that our method surpasses others in both classification accuracy and the minimization of selected genes.

Our paper paves the way for several innovative avenues in feature selection using meta-heuristic algorithms. We propose developing hybrid methods that combine the strengths of various approaches to tackle complex data challenges like high-dimensional and multi-label datasets. Future directions include focusing on model transparency and efficiency, adjusting algorithm parameters to data changes dynamically, and scaling to larger datasets. Additionally, integrating these algorithms with deep learning, establishing robust theoretical underpinnings, improving their explainability, and adapting them to specialized domains presents promising opportunities for advancement in this field.

Table 12 Comparison between the proposed method and other meta-heuristics in terms of accuracy (%) and the number of selected genes

Dataset	Criteria	Algorithm						
		PSO	GA	Firefly	GWO	ACO	WOA	**BHOA**
Lymphoma	Acc	100 ± 0	98.18 ± 1.61	98.48 ± 0.96	98.49 ± 0.96	97.72 ± 2.83	100 ± 0	**100 ± 0**
	#Gene	8.3 ± 0.36	9.1 ± 1.31	13.5 ± 1.26	9.16 ± 1.94	4.83 ± 1.32	5.42 ± 0.38	**2 ± 0**
	rank	(3)	(6)	(5)	(4)	(7)	(2)	**(1)**
Prostate	Acc	95.70 ± 1.7	97.43 ± 1.38	91.5 ± 0.71	96.58 ± 1.25	97.35 ± 1.92	94.78 ± 0.49	**97.72 ± 0.45**
	#Gene	14.3 ± 1.7	9.2 ± 1.06	26 ± 0.87	19.2 ± 1.38	18.42 ± 1.59	22.77 ± 1.68	**5.4 ± 1.26**
	rank	(5)	(2)	(7)	(4)	(3)	(6)	**(1)**
Brain-1	Acc	94.68 ± 1.49	88.11 ± 0.62	91.83 ± 0.62	86.47 ± 1.63	89.25 ± 3.56	92.75 ± 1.94	**96.45 ± 0.49**
	#Gene	15.36 ± 1.32	18.62 ± 1.83	21 ± 0.89	15.33 ± 1.75	16.33 ± 2.42	14.75 ± 1.23	**8.81 ± 1.25**
	rank	(2)	(6)	(4)	(7)	(5)	(3)	**(1)**
DLBCL	Acc	100 ± 0	98.46 ± 1.13	99.08 ± 1.05	99.29 ± 1.09	97.23 ± 1.69	100 ± 0	**100 ± 0**
	#Gene	6.57 ± 1.27	18.3 ± 1.02	13.33 ± 1.90	11.5 ± 2.34	5.83 ± 0.75	9.6 ± 0.24	**2.8 ± 0.63**
	rank	(2)	(6)	(5)	(4)	(7)	(3)	**(1)**
Colon	Acc	90.89 ± 1.58	86.79 ± 2.06	88.36 ± 1.26	88.49 ± 1.98	86.82 ± 1.58	91.67 ± 1.31	**98.48 ± 0.87**
	#Gene	12.43 ± 1.82	15.14 ± 1.89	13.26 ± 1.84	12.63 ± 1.79	9.33 ± 1.51	10.65 ± 1.26	**7.36 ± 1.61**
	rank	(3)	(7)	(5)	(4)	(6)	(2)	**(1)**
Leukemia	Acc	99.79 ± 0.43	97.16 ± 1.79	97.42 ± 0.95	99.89 ± 0.54	97.68 ± 1.13	100 ± 0	**100 ± 0**
	#Gene	12.4 ± 1.34	15.5 ± 0.64	8.5 ± 1.76	8.45 ± 1.75	4.83 ± 1.32	8.33 ± 0.43	**2.6 ± 0.51**
	rank	(4)	(7)	(6)	(3)	(5)	(2)	**(1)**
SRBCT	Acc	100 ± 0	97.30 ± 1.68	99.36 ± 0.70	100 ± 0	98.07 ± 2.18	100 ± 0	**100 ± 0**
	#Gene	13.2 ± 1.04	15.6 ± 1.93	16.33 ± 1.21	14.33 ± 0.51	8.8 ± 1.31	17 ± 1.01	**5.53 ± 1.05**
	rank	(2)	(7)	(5)	(3)	(6)	(4)	**(1)**
Lung	Acc	89.78 ± 1.94	94.26 ± 1.04	94.91 ± 0.82	94.19 ± 0.62	93.95 ± 1.63	96.78 ± 1.45	**98.66 ± 0.23**
	#Gene	21.4 ± 1.37	24 ± 1.01	18 ± 1.09	16.75 ± 0.95	16.33 ± 3.07	18.85 ± 0.89	**8.36 ± 2.20**
	rank	(7)	(4)	(3)	(5)	(6)	(2)	**(1)**
Ovarian	Acc	100 ± 0	100 ± 0	99.88 ± 0.17	100 ± 0	99.86 ± 0.18	100 ± 0	**100 ± 0**
	#Gene	5.6 ± 0.78	18.35 ± 0.61	11.5 ± 1.52	7.5 ± 0.54	3.6 ± 0.55	8.12 ± 0.68	**3 ± 0**
	Rank	(2)	(5)	(6)	(3)	(7)	(4)	**(1)**
MLL	Acc	98.64 ± 1.62	98.36 ± 0.95	99.52 ± 0.76	98.84 ± 1.62	99.10 ± 1.12	100 ± 0	**100 ± 0**
	#Gene	8.76 ± 1.04	10.84 ± 1.72	12.83 ± 2.23	13.66 ± 1.63	7.66 ± 1.03	10.44 ± 1.12	**4.1 ± 0.85**
	rank	(6)	(7)	(3)	(5)	(4)	(2)	**(1)**
Average rank		(3.6)	(5.7)	(4.9)	(4.2)	(5.6)	(3)	**(1)**

The bold values signify the best accuracy, number of genes, and rank achieved by the specific algorithm for each dataset

Table 13 Comparison between the proposed approach and other methods in terms of accuracy (%) and the number of selected genes

Dataset	Method	Acc (#Gene)	Ref	Dataset	Method	Acc (#Gene)	Refs.
Lymphoma	RMRMR-HBA	100 (8.13)	[64]	Leukemia	RMRMR-HBA	100 (4.07)	[64]
	BPSO-CGA	100 (196)	[65]		BPSO-CGA	100 (300)	[65]
	BIRSW	82.14 (10.3)	[66]		DBH	100 (4)	[67]
	IT-bBOA	94 (30)	[67]		TOPSIS-Jaya	100 (16.1)	[69]
	CFC-iBPSO	100 (24)	[68]		CFC-iBPSO	100 (4.3)	[68]
	Proposed	**100 (2)**			**Proposed**	**100 (2.6)**	
Prostate	IT-bBOA	96 (30)	[67]	SRBCT	RMRMR-HBA	100 (9.13)	[64]
	IBPSO	92.16 (*)	[70]		BPSO-CGA	100 (880)	[65]
	BFO	97.42 (29)	[61]		CFC-iBPSO	100 (34.1)	[68]
	DRF0-CGS	97.06 (*)	[62]		HSA-MB	99.57 (8.9)	[72]
	IWSSr	94.3 (10)	[71]		MBEGA	99.23 (60.7)	[73]
	Proposed	**97.72 (5.4)**			**Proposed**	**100 (5.53)**	
Brain-1	BFO	90.37 (25)	[61]	Lung	BFO	93.11 (39)	[61]
	BPSO_TS	95.89 (*)	[73]		DRF0-CGS	98.66 (17)	[62]
	IBPSO	94.44 (*)	[70]		PSO dICA	97.95 (25)	[75]
	DEFS	96.30 (30)	[63]		BDE-X Rankf	98.0 (3)	[76]
	IG-IBPSO	92.22 (115)	[74]		DLFCC	83.00 (*)	77[]
	Proposed	**96.45 (8.81)**			**Proposed**	**98.66 (8.36)**	
DLBCL	DBH	100 (4.05)	[69]	Ovarian	RMRMR-HBA	100 (3.07)	[64]
	BFO	98.99 (8)	[61]		CFC-iBPSO	100 (3.3)	[68]
	DRF0-CGS	94.67 (11)	[62]		HSA-MB	99.81 (5.73)	[72]
	IWSSr	94.73 (30)	[71]		DBH	100 (2.6)	[69]
	BDE-X Rankf	92.9 (3)	[76]		TOPSIS-Jaya	99.52 (18.5)	[78]
	Proposed	**100 (2.8)**			**Proposed**	**100 (3)**	
Colon	RMRMR-HBA	97.85 (12.27)	[64]	MLL	RMRMR-HBA	100 (8)	[64]
	BPSO-CGA	96.7 (214)	[65]		CFC-iBPSO	100 (30.08)	[68]
	DBH	97.02 (14.4)	[69]		HSA-MB	99.55 (6.6)	[72]
	IT-bBOA	86 (30)	[67]		DBH	100 (5.25)	[69]
	CFC-iBPSO	94.89 (4.2)	[68]		TOPSIS-Jaya	99.62 (12.9)	[78]
	Proposed	**98.48 (7.36)**			**Proposed**	**100 (4.1)**	

The bold values signify the best accuracy and number of genes achieved by the specific method for each dataset

Table 14 Average ranking of accuracy values between 7 algorithms on ten biomedical datasets by Friedman test

Algorithm	Ranking
PSO	2.9
GA	4.4
Firefly	3.9
GWO	3.2
ACO	4.3
WOA	1.9
Proposed	**1**

The bold value signifies the highest rank achieved by the specific algorithm

Table 15 Average ranking of selected genes between 7 algorithms on ten biomedical datasets by Friedman test

Algorithm	Ranking
PSO	3.9
GA	5.8
Firefly	6
GWO	4.6
ACO	2.5
WOA	4.2
Proposed	**1**

The bold value signifies the highest rank achieved by the specific algorithm

Table 16 Post Hoc comparison regarding accuracy values

Comparison	*Z*-value	*p*-value	Result
Proposed vs PSO	− 2.070197	0.038434	H_0 is rejected
Proposed vs GA	− 3.985129	0.000067	H_0 is rejected
Proposed vs Firefly	− 3.36407	0.000768	H_0 is rejected
Proposed vs GWO	− 2.587746	0.009661	H_0 is rejected
Proposed vs ACO	− 4.088638	0.000043	H_0 is rejected
Proposed vs WOA	− 0.931589	0.351549	H_0 is not rejected

Table 17 Post Hoc comparison regarding selected genes

Comparison	*Z*-value	*p*-value	Result
Proposed vs PSO	− 3.001785	0.002684015	H_0 is rejected
Proposed vs GA	− 4.968472	0.000000675	H_0 is rejected
Proposed vs Firefly	− 5.175492	0.000000227	H_0 is rejected
Proposed vs GWO	− 3.726354	0.000194269	H_0 is rejected
Proposed vs ACO	− 1.552648	0.120507400	H_0 is not rejected
Proposed vs WOA	− 3.312315	0.000925274	H_0 is rejected

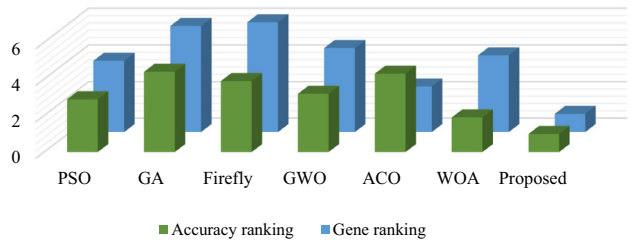

Fig. 14 Comparison between the ranking of existing algorithms and the proposed approach by the Friedman test

Acknowledgements The authors would like to thank anonymous reviewers for their time, valuable comments, constructive criticism, and suggestions, which greatly improved the paper.

Author contributions NM conceptualization, investigation, methodology, formal analysis, software, validation, writing. SPHB methodology, formal analysis, validation, writing. EP supervision.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare that this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Chizi, B., Rokach, L., Maimon, O.: A survey of feature selection techniques. In Encyclopedia of Data Warehousing and Mining, Second Edition, pp. 1888–1895. IGI Global (2009). https://doi.org/10.4018/978-1-60566-010-3.ch289

2. Shaffer, M., Borton, M.A., Bolduc, B., Faria, J.P., Flynn, R.M., Ghadermazi, P., Wrighton, K.C., et al.: kb_DRAM: annotation and metabolic profiling of genomes with DRAM in KBase. Bioinformatics **39**(4), btad110 (2023). https://doi.org/10.1093/bioinformatics/btad110
3. Wei, G., Zhao, J., Feng, Y., He, A., Yu, J.: A novel hybrid feature selection method based on dynamic feature importance. Appl. Soft Comput. **93**, 106337 (2020). https://doi.org/10.1016/j.asoc.2020.106337
4. Hammouri, A.I., Mafarja, M., Al-Betar, M.A., Awadallah, M.A., Abu-Doush, I.: An improved dragonfly algorithm for feature selection. Knowl.-Based Syst. **203**, 106131 (2020). https://doi.org/10.1016/j.knosys.2020.106131
5. Meraihi, Y., Gabis, A.B., Mirjalili, S., Ramdane-Cherif, A.: Grasshopper optimization algorithm: theory, variants, and applications. IEEE Access (2021). https://doi.org/10.1109/ACCESS.2021.3067597
6. Mehrabi, N., Pashaei, E. Application of horse herd optimization algorithm for medical problems. In: 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), pp. 1–6. IEEE (2021). https://doi.org/10.1109/INISTA52262.2021.9548366
7. Boroujeni, S.P.H., Pashaei, E.: Data clustering using chimp optimization algorithm. In: 2021 11th International Conference on Computer Engineering and Knowledge(ICCKE), pp. 296–301 IEEE (2021). https://doi.org/10.1109/ICCKE54056.2021.9721483
8. Mehrabi, N., Boroujeni, S.P.H.: Age estimation based on facial images using hybrid features and particle swarm optimization. In: 2021 11th International Conference on Computer Engineering and Knowledge (ICCKE), pp. 412–418. IEEE (2021). https://doi.org/10.1109/ICCKE54056.2021.9721496
9. Abualigah, L., Gandomi, A.H., Elaziz, M.A., Hamad, H.A., Omari, M., Alshinwan, M., Khasawneh, A.M.: Advances in meta-heuristic optimization algorithms in big data text clustering. Electronics **10**(2), 101 (2021). https://doi.org/10.3390/electronics10020101
10. Boroujeni, S.P.H., Razi, A.: IC-GAN: an improved conditional generative adversarial network for RGB-to-IR image translation with applications to forest fire monitoring. Expert Syst. Appl. **238**, 121962 (2024). https://doi.org/10.1016/j.eswa.2023.121962
11. Erdem, E., Bozkurt, F.: A comparison of various supervised machine learning techniques for prostate cancer prediction. Avrupa Bilim ve Teknol. Derg. **21**, 610–620 (2021). https://doi.org/10.31590/ejosat.802810
12. Kashef, S., Nezamabadi-pour, H.: A label-specific multi-label feature selection algorithm based on the Pareto dominance concept. Pattern Recogn. **88**, 654–667 (2019). https://doi.org/10.1016/j.patcog.2018.12.020
13. González, J., Ortega, J., Damas, M., Martín-Smith, P., Gan, J.Q.: A new multi-objective wrapper method for feature selection–accuracy and stability analysis for BCI. Neurocomputing **333**, 407–418 (2019). https://doi.org/10.1016/j.neucom.2019.01.017
14. Zhang, J., Luo, Z., Li, C., Zhou, C., Li, S.: Manifold regularized discriminative feature selection for multi-label learning. Pattern Recogn. **95**, 136–150 (2019). https://doi.org/10.1016/j.patcog.2019.06.003
15. Prabhakar, S.K., Lee, S.W.: Transformation based tri-level feature selection approach using wavelets and swarm computing for prostate cancer classification. IEEE Access (2020). https://doi.org/10.1109/ACCESS.2020.3006197
16. Le, T.T., Urbanowicz, R.J., Moore, J.H., McKinney, B.A.: STatistical Inference Relief (STIR) feature selection. Bioinformatics (2019). https://doi.org/10.1093/bioinformatics/bty788
17. Omuya, E.O., Okeyo, G.O., Kimwele, M.W.: Feature selection for classification using principal component analysis and information gain. Expert Syst. Appl. **174**, 114765 (2021). https://doi.org/10.1016/j.eswa.2021.114765
18. Bahassine, S., Madani, A., Al-Sarem, M., Kissi, M.: Feature selection using an improved Chi-square for Arabic text classification. J. King Saud Univ.-Comput. Inf. Sci. **32**(2), 225–231 (2020). https://doi.org/10.1016/j.jksuci.2018.05.010
19. Pashaei, E., Aydin, N.: Binary black hole algorithm for feature selection and classification on biological data. Appl. Soft Comput. **56**, 94–106 (2017). https://doi.org/10.1016/j.asoc.2017.03.002
20. de Souza, R.C.T., de Macedo, C.A., dos Santos Coelho, L., Pierezan, J., Mariani, V.C.: Binary coyote optimization algorithm for feature selection. Pattern Recogn. **107**, 107470 (2020). https://doi.org/10.1016/j.patcog.2020.107470
21. Sarlak, A., Razi, A., Chen, X., Amin, R.: Diversity maximized scheduling in roadside units for traffic monitoring applications. In: 2023 IEEE 48th Conference on Local Computer Networks (LCN), pp. 1–4. IEEE (2023). https://doi.org/10.1109/LCN58197.2023.10223373
22. Mostafa Bozorgi, S., Yazdani, S.: IWOA: an improved whale optimization algorithm for optimization problems. J. Comput. Design Eng. **6**(3), 243–259 (2019). https://doi.org/10.1016/j.jcde.2019.02.002
23. Khishe, M., Mosavi, M.R.: Chimp optimization algorithm. Expert Syst. Appl. **149**, 113338 (2020). https://doi.org/10.1016/j.eswa.2020.113338
24. Azizyan, G., Miarnaeimi, F., Rashki, M., Shabakhty, N.: Flying Squirrel Optimizer (FSO): a novel SI-based optimization algorithm for engineering problems. Iran. J. Optimiz. **11**(2), 177–205 (2019).
25. Pierezan, J., Dos Santos Coelho, L.: Coyote optimization algorithm: a new metaheuristic for global optimization problems. In: 2018 IEEE Congress on Evolutionary Computation, CEC 2018—Proceedings (2018). https://doi.org/10.1109/CEC.2018.8477769
26. Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H.: Harris hawks optimization: algorithm and applications. Futur. Gener. Comput. Syst. **97**, 849–872 (2019). https://doi.org/10.1016/j.future.2019.02.028
27. Mirjalili, S.Z., Mirjalili, S., Saremi, S., Faris, H., Aljarah, I.: Grasshopper optimization algorithm for multi-objective optimization problems. Appl. Intell. **48**, 805–820 (2018). https://doi.org/10.1007/s10489-017-1019-8
28. Pashaei, E., Pashaei, E.: Gene selection using intelligent dynamic genetic algorithm and random forest. In: ELECO 2019—11th International Conference on Electrical and Electronics Engineering (2019). https://doi.org/10.23919/ELECO47770.2019.8990557
29. MiarNaeimi, F., Azizyan, G., Rashki, M.: Horse herd optimization algorithm: A nature-inspired algorithm for high-dimensional optimization problems. Knowl.-Based Syst. **213**, 106711 (2021). https://doi.org/10.1016/j.knosys.2020.106711
30. Boroujeni, S.P.H., Razi, A., Khoshdel, S., Afghah, F., Coen, J.L., ONeill, L., Vamvoudakis, K.G. et al.: A Comprehensive Survey of Research Towards AI-Enabled Unmanned Aerial Systems in Pre-, Active-, and Post-Wildfire Management. Springer, New York (2024). https://doi.org/10.48550/arXiv.2401.02456
31. McDonnell, S.M., Poulin, A.: The equid ethogram: a practical field guide to horse behavior—Sue M. McDonnell. Appl. Anim. Behav. Sci. **2003**, 789 (2003)
32. Levine, M.A.: Domestication and early history of the horse. In: The Domestic Horse: The Evolution, Development, and Management of its Behaviour, pp. 5–22 Springer, New York (2005)
33. Ghosh, K.K., Singh, P.K., Hong, J., Geem, Z.W., Sarkar, R.: Binary social mimic optimization algorithm with x-shaped transfer function for feature selection. IEEE Access **8**, 97890–97906 (2020). https://doi.org/10.1109/ACCESS.2020.2996611
34. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1226–1238 (2005). https://doi.org/10.1109/TPAMI.2005.159

35. Holland, J.H.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT press, Cambridge (1992)
36. Van Laarhoven, P.J., Aarts, E.H., van Laarhoven, P.J., Aarts, E.H.: Simulated Annealing, pp. 7–15. Springer, Netherlands (1987). https://doi.org/10.1007/978-94-015-7744-1_2
37. Araujo, L.: Genetic programming for natural language processing. Genet. Program Evolvable Mach. **21**(1–2), 11–32 (2020). https://doi.org/10.1007/s10710-019-09361-5
38. Knowles, J.D., Corne, D.W.: M-PAES: a memetic algorithm for multiobjective optimization. In: Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512), vol. 1, pp. 325–332. IEEE (2000). https://doi.org/10.1109/CEC.2000.870313
39. Kuo, R.J., Zulvia, F.E.: The gradient evolution algorithm: a new metaheuristic. Inf. Sci. **316**, 246–265 (2015). https://doi.org/10.1016/j.ins.2015.04.031
40. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. Adv. Eng. Softw. **69**, 46–61 (2014). https://doi.org/10.1016/j.advengsoft.2013.12.007
41. Kaveh, A., Ilchi Ghazaan, M.: A new meta-heuristic algorithm: vibrating particles system. Sci. Iran. **24**(2), 551–566 (2017). https://doi.org/10.24200/sci.2017.2417
42. Al-Madi, N., Faris, H., Mirjalili, S.: Binary multi-verse optimization algorithm for global optimization and discrete problems. Int. J. Mach. Learn. Cybern. **10**, 3445–3465 (2019). https://doi.org/10.1007/s13042-019-00931-8
43. Boroujeni, S.P.H., Pashaei, E.: A novel hybrid gene selection based on random forest approach and binary dragonfly algorithm. In: 2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–8. IEEE (2021). https://doi.org/10.1109/CCE53527.2021.9633105
44. Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: BGSA: binary gravitational search algorithm. Nat. Comput. **9**, 727–745 (2010). https://doi.org/10.1007/s11047-009-9175-3
45. Varaee, H., Ghasemi, M.R.: Engineering optimization based on ideal gas molecular movement algorithm. Eng. Comput. **33**, 71–93 (2017). https://doi.org/10.1007/s00366-016-0457-y
46. Rashedi, E., Nezamabadi-Pour, H., Saryazdi, S.: GSA: a gravitational search algorithm. Inf. Sci. **179**(13), 2232–2248 (2009). https://doi.org/10.1016/j.ins.2009.03.004
47. Dehghani, M., Montazeri, Z., Trojovská, E., Trojovský, P.: Coati Optimization Algorithm: a new bio-inspired metaheuristic algorithm for solving optimization problems. Knowl.-Based Syst. **259**, 110011 (2023). https://doi.org/10.1016/j.knosys.2022.110011
48. Li, A.D., Xue, B., Zhang, M.: Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies. Appl. Soft Comput. **106**, 107302 (2021). https://doi.org/10.1016/j.asoc.2021.107302
49. Qasim, O.S., Algamal, Z.Y.: Feature selection using different transfer functions for binary bat algorithm. Int. J. Math. Eng. Manage. Sci. **5**(4), 697 (2020). https://doi.org/10.33889/IJMEMS.2020.5.4.056
50. Tawhid, M.A., Ibrahim, A.M.: Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. Int. J. Mach. Learn. Cybern. **11**, 573–602 (2020). https://doi.org/10.1007/s13042-019-00996-5
51. Zhang, Y., Liu, R., Wang, X., Chen, H., Li, C.: Boosted binary Harris hawks optimizer and feature selection. Eng. Comput. **37**, 3741–3770 (2021). https://doi.org/10.1007/s00366-020-01028-5
52. Dhiman, G., Oliva, D., Kaur, A., Singh, K.K., Vimal, S., Sharma, A., Cengiz, K.: BEPO: A novel binary emperor penguin optimizer for automatic feature selection. Knowl.-Based Syst. **211**, 106560 (2021). https://doi.org/10.1016/j.knosys.2020.106560
53. Hu, P., Pan, J.S., Chu, S.C.: Improved binary grey wolf optimizer and its application for feature selection. Knowl.-Based Syst. **195**, 105746 (2020). https://doi.org/10.1016/j.knosys.2020.105746
54. Lori, A.A.R.: Optimal path planning for aerial load transportation in complex environments using PSO-improved artificial potential fields. arXiv (2023). https://doi.org/10.48550/arXiv.2311.10675
55. Lee, K.Y., Park, J.B.: Application of particle swarm optimization to economic dispatch problem: advantages and disadvantages. In: 2006 IEEE PES power systems conference and exposition (pp. 188–192). IEEE (2006). https://doi.org/10.1109/PSCE.2006.296295
56. Haeri Boroujeni, S.P., Pashaei, E.: A hybrid chimp optimization algorithm and generalized normal distribution algorithm with opposition-based learning strategy for solving data clustering problems. Iran J. Comput. Sci. **2023**, 1–37 (2023). https://doi.org/10.1007/s42044-023-00160-x
57. Dhiman, G., Kumar, V.: Emperor penguin optimizer: a bio-inspired algorithm for engineering problems. Knowl.-Based Syst. **159**, 20–50 (2018). https://doi.org/10.1016/j.knosys.2018.06.001
58. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinf. Comput. Biol. **3**(02), 185–205 (2005). https://doi.org/10.1142/S0219720005001004
59. Mirjalili, S., Lewis, A.: S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm Evol. Comput. **9**, 1–14 (2013). https://doi.org/10.1016/j.swevo.2012.09.002
60. Mafarja, M., Mirjalili, S.: Whale optimization approaches for wrapper feature selection. Appl. Soft Comput. **62**, 441–453 (2018). https://doi.org/10.1016/j.asoc.2017.11.006
61. Wang, H., Jing, X., Niu, B.: A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data. Knowl.-Based Syst. **126**, 8–19 (2017). https://doi.org/10.1016/j.knosys.2017.04.004
62. Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A.: Distributed feature selection: an application to microarray data classification. Appl. Soft Comput. **30**, 136–150 (2015). https://doi.org/10.1016/j.asoc.2015.01.035
63. Khushaba, R.N., Al-Ani, A., Al-Jumaily, A.: Feature subset selection using differential evolution and a statistical repair mechanism. Expert Syst. Appl. **38**(9), 11515–11526 (2011). https://doi.org/10.1016/j.eswa.2011.03.028
64. Alomari, O.A., Khader, A.T., Al-Betar, M.A., Awadallah, M.A.: A novel gene selection method using modified MRMR and hybrid bat-inspired algorithm with β-hill climbing. Appl. Intell. **48**, 4429–4447 (2018). https://doi.org/10.1007/s10489-018-1207-1
65. Chuang, L.Y., Yang, C.H., Li, J.C., Yang, C.H.: A hybrid BPSO-CGA approach for gene selection and classification of microarray data. J. Comput. Biol. **19**(1), 68–82 (2012). https://doi.org/10.1089/cmb.2010.0064
66. Ruiz, R., Riquelme, J.C., Aguilar-Ruiz, J.S.: Incremental wrapper-based gene selection from microarray data for cancer classification. Pattern Recogn. **39**(12), 2383–2392 (2006). https://doi.org/10.1016/j.patcog.2005.11.001
67. Sadeghian, Z., Akbari, E., Nematzadeh, H.: A hybrid feature selection method based on information theory and binary butterfly optimization algorithm. Eng. Appl. Artif. Intell. **97**, 104079 (2021). https://doi.org/10.1016/j.engappai.2020.104079
68. Jain, I., Jain, V.K., Jain, R.: Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. Appl. Soft Comput. **62**, 203–215 (2018). https://doi.org/10.1016/j.asoc.2017.09.038
69. Pashaei, E., Pashaei, E.: Gene selection using hybrid dragonfly black hole algorithm: A case study on RNA-seq COVID-19 data. Anal. Biochem. **627**, 114242 (2021). https://doi.org/10.1016/j.ab.2021.114242
70. Chuang, L.Y., Chang, H.W., Tu, C.J., Yang, C.H.: Improved binary PSO for feature selection using gene expression data. Comput.

Biol. Chem. **32**(1), 29–38 (2008). https://doi.org/10.1016/j.compbiolchem.2007.09.005
71. Wang, A., An, N., Chen, G., Li, L., Alterovitz, G.: Accelerating wrapper-based feature selection with K-nearest-neighbor. Knowl.-Based Syst. **83**, 81–91 (2015). https://doi.org/10.1016/j.knosys.2015.03.009
72. Shreem, S.S., Abdullah, S., Nazri, M.Z.A.: Hybridising harmony search with a Markov blanket for gene selection problems. Inf. Sci. **258**, 108–121 (2014). https://doi.org/10.1016/j.ins.2013.10.012
73. Zhu, Z., Ong, Y.S., Dash, M.: Markov blanket-embedded genetic algorithm for gene selection. Pattern Recogn. **40**(11), 3236–3248 (2007). https://doi.org/10.1016/j.patcog.2007.02.007
74. Chuang, L.Y., Ke, C.H., Yang, C.H.: A hybrid both filter and wrapper feature selection method for microarray classification. arXiv:1612.08669 (2016)
75. Mollaee, M., Moattar, M.H.: A novel feature extraction approach based on ensemble feature selection and modified discriminant independent component analysis for microarray data classification. Biocybernet. Biomed. Eng. **36**(3), 521–529 (2016). https://doi.org/10.1016/j.bbe.2016.05.001
76. Apolloni, J., Leguizamón, G., Alba, E.: Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. Appl. Soft Comput. **38**, 922–932 (2016). https://doi.org/10.1016/j.asoc.2015.10.037
77. Sharma, A., Rani, R.: An optimized framework for cancer classification using deep learning and genetic algorithm. J. Med. Imaging Health Inf. **7**(8), 1851–1856 (2017). https://doi.org/10.1166/jmihi.2017.2266
78. Chaudhuri, A., Sahu, T.P.: A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. Comput. Electr. Eng. **90**, 106963 (2021). https://doi.org/10.1016/j.compeleceng.2020.106963
79. Conover, W.J., Iman, R.L.: Rank transformations as a bridge between parametric and nonparametric statistics. Am. Stat. **1981**, 124–129 (1981). https://doi.org/10.1080/00031305.1981.10479327