

An enhanced grey wolf optimization method for feature selection and explainable prediction in chronic disease analytics

Sibasish Dhibar ^a, Deepa Gupta ^{b,*}, Gopalakrishnan E.A. ^{a,b}, Susmitha Vekkot ^c

^a Department of Artificial Intelligence, Amrita School of Artificial Intelligence, Bengaluru, Amrita Vishwa Vidyapeetham, India

^b Department of Computer Science and Engineering, Amrita School of Computing Bengaluru, Amrita Vishwa Vidyapeetham, India

^c Department of Electronics and Communication Engineering, Amrita School of Engineering Bengaluru, Amrita Vishwa Vidyapeetham, India



ARTICLE INFO

Dataset link: <https://jundongli.github.io/scikit-feature/datasets.html>, <http://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>, <http://archive.ics.uci.edu/dataset/46/hepatitis>, <https://www.kaggle.com/datasets/malchi/diabetes-data-set>

Keywords:

Metaheuristic algorithms
Median random wondering fitness
Grey wolf optimization
Feature selection
Chronic disease analytics
Shapley Additive Explanations

ABSTRACT

The advent of artificial intelligence paradigms has contributed a great deal towards improving predictive accuracy in medicine especially in the prompt diagnosis of chronic diseases. Feature selection is crucial in refining classification models by identifying and eliminating non-informative or repetitive data attributes. In the process of feature selection, the classical Grey Wolf Optimizer (GWO) suffers from low convergence efficiency and likelihood of stagnation at local optima. This research explores the median random wandering fitness-based Grey Wolf Optimization (MRWF-GWO) method to enhance local search performance and obtain an optimal balance between exploration and exploitation. The effectiveness of the proposed approach was validated through a systematic analysis using classification metrics across ten chronic disease datasets. Additionally, the performance of MRWF-GWO is compared with several state-of-the-art metaheuristic algorithms. The experimental analysis substantiates that MRWF-GWO consistently achieves superior performance in terms of search efficiency, classification accuracy, feature subset size, stability, computational time, and convergence rate. MRWF-GWO yielded an average accuracy of 92.89% across the ten chronic disease datasets considered in the study. The MRWF-GWO algorithm achieved a substantial feature size reduction (95%–99%) compared to the original set thereby demonstrating consistent efficiency in feature selection. Additionally for Leukemia and Lung datasets, the proposed MRWF-GWO method has the fastest execution time of 4.33s and 6.03s compared to all other metaheuristic optimization algorithms (MHA). The outcome of the predictive model is used to comprehend the most relevant clinical features that contributed to predicting chronic diseases, using an explainable machine learning technique known as Shapley Additive Explanations.

1. Introduction

The quicker evolution of healthcare methodologies and the exponential growth of medical data have guided personalized medicine where treatment plans are designed to individual patients and medical histories. Healthcare professionals rely on these insights to make informed decisions thereby necessitating a clear understanding of how models derive their predictions. Retrieving essential information from large volumes of data poses a significant challenge for humans [1]. In machine learning (ML), it is anticipated that healthcare applications could be developed to perform certain tasks with equal or greater accuracy than human physicians [2,3]. Over the past several decades, vast amounts of data have been generated across diverse domains, including biology [4], industrial production [5], and image recognition [6].

While having a large dataset can be useful, the quality of the data is more significant to make informed predictions [7]. It is imperative that clean, accurate and relevant data leads to better model performance.

Noisy, biased or incomplete data can mislead the model which in turn leads to reduced model performance. Hashemi et al. [8] discussed that the presence of irrelevant features lengthens the training time and makes the feature selection harder due to the curse of dimensionality. Due to these constraints, selecting features from large medical datasets is a pivotal step in developing efficient and accurate predictive models in healthcare analytics. A substantial number of samples is required to ensure effective model generalization as a small feature set can lead to underfitting. Consequently, feature selection (FS) is employed as a preprocessing approach to extract the most relevant attributes for the learning model. It has been widely applied to address diverse challenges across various domains including biological data processing [9], bioinformatics [10], medical [11], drug design [12], and so on.

FS may be further divided into filter, embedded, and wrapper based approaches based on integration with learning models [13]. Nevertheless, including numerous non-informative features can introduce the curse of dimensionality making learning and generalization

* Corresponding author.

E-mail address: g.deepa@blr.amrita.edu (D. Gupta).

more difficult [14]. Dimensionality reduction involves optimizing the dataset by finding a matrix with fewer columns to make it more manageable and effective. The combination of dimensionality reduction and optimization has emerged as a vital research area in numerous fields including predictive modeling applications in healthcare analytics. FS serves as an essential preprocessing stage to minimize data size and discard redundant or irrelevant features. In recent decades, several metaheuristic algorithms (MHAs) have been applied to efficiently solve FS-related optimization tasks. Several efficient MHAs have been developed for FS such as the two-archive multi-objective artificial bee colony algorithm (TMABC-FS) [15], binary differential evolution with self-learning (MOFS-BDE) [16], variable-size cooperate [17] and return-cost-based binary FFA (RC-BBFA) [18]. Despite enhancements to MHAs, challenges such as slow convergence, getting stuck in local optima and accuracy trade-off still exist. This research is inspired by the fundamental principle of the No Free Lunch (NFL) Theorem [19] which states that it is not feasible for a single metaheuristic optimization algorithm to achieve optimal performance across all optimization tasks. Nevertheless, considerable interest remains in exploring how MHAs can enhance classification performance particularly for large-scale datasets.

We consider grey wolf optimization (GWO) for feature selection from chronic medical datasets. The drawbacks of this algorithm include slow convergence, decreased model accuracy and the possibility of being trapped in local optimum. As a result, there is only a minor improvement in the algorithm's ability to converge. Since the results of the standard binary GWO (BGWO) algorithm are not acceptable for optimization of FS [20], modified variants of GWO have been used. Researchers in [21] proposed an enhanced binary GWO algorithm. Authors in [22] introduced a two-stage mutation-based GWO to enhance the search mechanism. Researchers in [23] hybridized GWO with particle swarm optimization (PSO) for efficient feature selection. Wang et al. [24] proposed a wrapper-based method using a KNN distance matrix for evaluating feature subsets. Ay et al. [25] integrated MHAs with KNN to improve heart disease diagnosis and survival prediction. Nssibi et al. [26] analyzed the effectiveness of MHAs and KNN for feature selection in medical datasets.

As artificial intelligence (AI) systems evolve and become increasingly intricate and advanced, the imperative for these systems to be transparent and understandable has risen to the forefront. Explainable AI (XAI) methods strive to enhance the clarity and reliability of AI models by elucidating their decision-making frameworks. The essence of XAI strategy lies in improving the interpretability and transparency of AI systems applied in medical diagnosis [27,28] thereby narrowing the divide between cutting-edge AI innovations and their practical implementation in healthcare domains. Shapley Additive Explanations (SHAP) [29] is recognized as a pivotal XAI technique instrumental in making the predictions of complex AI models transparent. The motivation for using SHAP is to provide local and understandable explanations for AI model predictions. SHAP is a widely recognized XAI technique applied to the detection of chronic diseases using smart healthcare approaches.

1.1. Novelty and research contributions

This work presents a novel advancement in the GWO algorithm by integrating median wandering random fitness values, substantially boosting its exploration ability and overall optimization performance. This aims to improve the convergence speed and ability to escape local optima of the standard GWO. The proposed approaches are applied to a KNN classifier and evaluated using ten medical benchmark datasets sourced from GitHub and the UCI Machine Learning Repository. The performance of the proposed approaches is compared with the state of the art. The main contributions of this research work are summarized below:

- The introduction of the Median Random Wandering Fitness-based Grey Wolf Optimizer (MRWF-GWO) substantially improves the

GWO mechanism by refining the balance between exploration and exploitation across the solution space.

- Performance evaluations of the MRWF-GWO algorithm with respect to other FS algorithms including differential evaluation (DE), fruit fly optimization algorithm (FOA), PSO, sine cosine algorithm (SCA), and whale optimization algorithm (WOA).
- The performance of the proposed method is benchmarked against other advanced techniques on a chronic disease dataset, with an objective function that jointly minimizes feature selection and classification error.
- Extensive quantitative assessments are conducted in terms of accuracy, average fitness values, F1-score, standard deviation, and number of selected features.
- XAI is incorporated using SHAP to analyze and improve the decision making in models.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature. Section 3 introduces the foundational concepts of GWO and its schematic representation. Section 4 details the evaluation metrics and experimental design. Section 5 discusses the results and their analysis. Section 6 concludes the work, highlighting the key insights and suggesting avenues for future investigation.

2. Related work

Several studies have explored enhancements to MHAs by focusing on tuning control parameters, modifying the position update rules or integrating GWO with other techniques. The rapid increase in feature selection for large amounts of data in healthcare has led researchers to develop more advanced and reliable classification models. Table 1 summarizes a comparison among the relevant research with various datasets including chronic medical datasets. Filter, wrapper, and hybrid approaches are the prominent methods discussed. Additionally, Table 1 shows that most researchers have applied KNN and SVM for classification problems using different datasets. Ephzibah [30] employed a fuzzy genetic algorithm for feature selection, effectively reducing the feature set and identifying features with significant impact on diagnosis through fuzzy rules. Sahoo & Chandra [31] used non-dominated sorting grey wolf optimizer (NSGWO) to enhance the categorization of cervical lesions by reducing the number of textural characteristics and increasing classification accuracy. Mafarja and Mirjalili [33] investigated simulated annealing with WOA (WOA-SA) for feature extraction. Various variants of the PSO technique have been improved by Thara and Gunasundari [32], Faris, Mafarja, et al. [34] and Ji et al. [37] to optimize feature reduction. All these studies employ KNN classification across multiple datasets and used diverse performance metrics including accuracy, elapsed time and fitness values.

Enreddy [36] introduced a modified version of GWO by incorporating random wolf positioning to enhance the optimization of the fitness function and presented statistical analysis. Feature classification was performed using KNN, SVM, and XGBoost algorithms. Hou et al. [35] presented the BIFFOA, a binary fruit fly algorithm based on an S-shaped transfer function. However, this algorithm exhibits certain limitations in the aspect of dimensionality reduction. Zhang et al. [39] proposed a Gaussian mutational chaotic FOA algorithm (MCFOA) for feature selection combined with a KNN classifier across 20 different datasets. Hans and Kaur [38] designed an opposition-based enhancement of the GWO (OEGWO) for feature selection in medical datasets. However, they considered only one medical dataset and obtained fitness values, accuracy, and feature size. Wang and Chen [40] suggested an enhanced whale optimization algorithm (WOA) that integrates several methodologies for conducting parameter optimization and feature selection concurrently. Recently, Raveenthini et al. [55] proposed a unified framework for multiocular disease detection. The framework incorporated non-linear features to classify AMD, cataract, diabetic retinopathy (DR), glaucoma and healthy cases. Their approach leveraged a hybrid optimization algorithm (JA-HHO) for feature selection and XGBoost (XGB)

Table 1

Summary of feature selection methods using existing MHAs [Abbreviations: GA Genetic Algorithm; PSO Particle Swarm Optimization; WOA Whale Optimization algorithm; FOA Fruit Fry Algorithm; SCA Sine Cosine Algorithm; DE Differential Evaluation; CSA: Clone Selection Algorithm; GJO Golden Jackal Optimization; EA: Evolutionary Algorithm; XGB: XGBoost].

Author	Methods	Contribution	Classifier	Assessment metric	# of datasets
Ephzibah [30]	GA	GA and Fuzzy Logic (Fuzzy-GA)	Fuzzy rule	Fitness, Accuracy	1, UCI
Sahoo and Chandra [31]	GWO	Multi Objective Grey Wolf (NGWO)	K-NN	Fitness, Accuracy	2, UCI
Tharu and Gunasundari [32]	PSO	Adaptive Feature Select Particle Optimizer (AFPSO)	K-NN	Elapsed time, Accuracy	1, UCI
Mafarja and Mirjalili [33]	WOA	Amalgamation With SA (WOASA)	K-NN	Fitness, Features, Accuracy	18, UCI
Faris, Mafarja, et al. [34]	PSO	Amalgamation with Adaptive Strategies (MOPSOASFS)	KNN	Fitness, Elapsed time, Features, Accuracy	14, UCI
Hou et al. [35]	FOA	Binary Improved of FOA (BIFFOA)	KNN	Elapsed time F1-score, Features, Accuracy	25, UCI
Enireddy [36]	GWO	Binary Grey Wolf (BGWO)	K-NN, SVM, XGB	Accuracy, F1-Score	1, UCI
Ji et al. [37]	PSO	Mutation Operation and Levy Flight (IBPSO)	KNN	Fitness, Elapsed time, Features, Accuracy	16, UCI
Y. Zhang et al. [16]	DE	Self-Learning and Binary Mutation Operation (MOFSBDE)	KNN	Error, Elapsed time, Hyper-volume, rate, Features	20, UCI
Hans and Kaur [38]	GWO	Opposition-Based Enhanced Grey Wolf Optimization (OBGWO)	SVM	Fitness, Accuracy, Features	1, UCI
X. Zhang et al. [39]	FOA	Chaotic Local Search and the Gaussian Mutation Factor (MCFOA)	K-NN	Error rate, Features,	7, UCI
M. Wang and Chen [40]	WOA	Chaotic Multiswarm Whale Optimizer (CMWOA)	SVM	Elapsed time, Features, Accuracy	2, UCI
Nadimi-Shahraki et al. [41]	GWO	Improved GWO (I-GWO)	–	Fitness, Friedman test, MAE	20, CEC 2018
Jiang et al. [42]	GWO	Diversity enhanced Strategy GWO (DSGWO)	–	Fitness, Wilcoxon rank-sum test, Elapsed time	16, IEEE CEC 2014
Sahoo et al. [43]	GWO	Non-dominated Sorting based GWO (NSGWO)	SVM	Accuracy, F1, Fitness, Features	5, http://www.gems-system.org
J. Wang et al. [44]	GWO	Adaptive Balanced GWO (ABGWO)	KNN	Accuracy, Feature size, Elapsed time	12, Arizona State University,
Nadimi-Shahraki et al. [45]	WOA	Pooling Mechanism and Three Different (BE-WOA) Search Strategies	K-NN	University of California Irvine Accuracy, Sensitivity,	10, UCI
Kale and Yuzgec [46]	SCA	Improve With Various Updating Strategies (ImpSCAs)	K-NN	Specificity	10, UCI
Ramasamy Rajammal et al. [47]	GWO	Improve With Mutation Operator (BIGWO)	K-NN	Fitness, Features, Accuracy	4, UCI
Preeti and Deep [48]	GWO	The Random Walk and Dispersion Factor (RWGWO)	K-NN	Fitness, Features, Accuracy, Elapsed time, Fitness, Features, Accuracy	18, UCI, GitHub
Z. Li [49]	GWO	Opposition-Learning Golden-Sine and GWO (OGGWGO)	K-NN	Elapsed time, Fitness, Features, Accuracy	18, UCI, GitHub
Ou et al. [50]	GWO	Clone Selection Algorithm and GWO (pGWO-CSA)	–	Fitness, Wilcoxon test, Elapsed time	15, CEC 2018
Askr et al. [51]	GJO	Binary Enhanced GJO (BEGJO)	KNN	Elapsed time, Fitness, Features, Accuracy	15, UCI, Arizona State University
Sun et al. [52]	EA	Correlation-Redundancy Guided EA (CRGEA)	KNN	Elapsed time, Fitness, Features, Accuracy	16, UCI
Huang et al. [53]	MIGWO	Multi-strategy Improved Grey Wolf Optimizer	KNN	Fitness, Features, Accuracy	10, UCI
Ihsan et al. [54]	MGWO-MCI	Multi-solution cross over integration	KNN	Fitness, Features, Accuracy	18, UCI, NSL-KDD

for classification. Maazalahi and Hosseini [56] introduced a hybrid approach combining the SailFish Optimizer (SFO), Whale Optimization Algorithm (WOA), PSO and K-means for attack detection. Sadeghian et al. [57] conducted a comprehensive review of various meta-heuristic feature selection methods developed and implemented across diverse datasets to address feature selection challenges effectively.

Nadimi-Shahraki et al. [45] introduced improved GWO (I-GWO) evaluated using the Friedman test across 20 benchmarks from the CEC 2018 dataset. Meanwhile, Jiang et al. [42] developed a diversity enhanced strategy GWO (DSGWO) assessed through fitness scores, the Wilcoxon rank-sum test and elapsed time metrics on 16 benchmarks from IEEE CEC. These studies reflect ongoing advancements in refining GWO by combining it with various strategies to boost its efficiency and reliability in optimization. To address the imbalance between global exploration and local exploitation, various researchers [43,44] have implemented the standard GWO algorithm. They aimed to obtain the best fitness values for feature reduction and achieved significant feature contributions using SVM and KNN classification respectively. E-GWO was proposed by researchers in [45] using a pooling mechanism in the standard GWO. They also used KNN classification for ten UCI datasets. Alternatively, four variants of the boosting SCA were studied by Kale & Yüzgeç [46] for feature selection and global optimization. Improvements were made to the standard GWO by Rajammal et al. [47] and Preeti and Deep [48] by incorporating a mutation operator and dispersion factor respectively. Both enhancements were applied to multiple datasets using KNN classification thereby retaining key features and optimizing feature reduction from different datasets. A hybrid grey wolf optimizer incorporating a clone selection algorithm (pGWO-CSA) [50] and a golden sine algorithm [49] were used on various chronic disease datasets.

Askr et al. [51] proposed binary enhanced golden jackal optimization (BEGJO) using KNN classifier. Moreover, statistical evaluations were conducted to further evaluate the proposed algorithm. Sun et al. [52] introduced a correlation-redundancy guided evolutionary algorithm (CRGEA) for high-dimensional feature selection, aiming to simultaneously optimize classification accuracy while maintaining a minimum selected feature set. Huang et al. [53] proposed the

Multi-strategy Improved GWO (MIGWO) to address high-dimensional feature selection challenges. Experimental evaluations on ten high-dimensional datasets demonstrated that MIGWO effectively reduces the subset feature size while achieving higher classification accuracy. Khaseeb et al. [58] developed an approach that balances exploration and exploitation within the search space by integrating two metaheuristic algorithms: BGWO in even iterations and Binary PSO (BPSO) in odd iterations. Ewees et al. [59] proposed an improved GWO termed as MGWO-MCI that enhances performance by integrating GWO's hunting strategy with multi-solution crossover.

Optimization algorithms and its variants have significantly evolved to address complex optimization issues like FS across various research areas Classical GWO methods are known to converge prematurely [60], prompting researchers to develop improvements to achieve an improved balance between global exploration and local exploitation. The comprehensive summary in Table 1 highlights that research utilizing MHAs has led to significant improvements in fitness, accuracy and processing time, underscoring the effectiveness of these methods in handling complex datasets in healthcare and other domains. Additionally, the use of hybrid and ensemble methods has increased due to the rapid growth of data in healthcare. The widespread application of KNN and SVM classifiers in these studies highlights the effectiveness of MHAs in managing large and complex datasets. Ongoing advancements in GWO and other MHAs are proving to be crucial in enhancing feature selection techniques, indicating a promising area of active research.

MHAs have been employed in various real world applications ranging from medical and cybersecurity space. GWO based feature selection is successfully applied for cyber threat detection in wireless communication networks in [61]. The research combines Binary Gravitational Search Algorithm (BGSA) and BGWO for feature selection. The framework significantly improves detection performance employing classifiers like Random Forest and AdaBoost. Using the UNSW-NB15 dataset, it achieves high accuracy (99.41%) and a low false positive rate (0.03%) with only 4 selected features. Among medical research paradigms, [62] presented a hybrid deep neural network model for predicting Coronary Heart Disease (CHD) using the BRFSS-2015 dataset. Through optimal feature selection using correlation-based scores and

class balancing via a cluster-enriched strategy, the model attained a classification accuracy of 98.28% after fine-tuning the BiLSTM and GRU architectures with randomized search cross-validation. Along similar lines, the study in [63] explores the use of PSO and other MHAs viz. Genetic Algorithm (GA), Simulated Annealing (SA), Ant Colony Optimization (ACO), and Tabu Search (TS) for detecting muscle activation in surface electromyography (sEMG) signals. PSO outperforms other algorithms in median accuracy, F1-score and speed but shows lower stability than GA and ACO, emphasizing the importance of cost function design in optimizing performance.

Despite advancements such as randomized initialization, hybrid methods and other improvements to GWO the algorithm still faces challenges in convergence speed and solution quality and offers scope for better computational efficiency. Eventhough a variety of feature selection methods have been proposed, the use of median random wandering fitness in GWO-based subset selection is rare. The development of a feature selection approach utilizing the GWO of incremental usefulness is still an open issue. We developed a new variant of GWO, namely MRWF-GWO. The GWO variant is employed in three separate ways to select optimal features across multiple datasets.

- A nonlinear function is employed to modulate the iterative decay of the GWO convergence factor for maintaining a stochastic balance between exploration and exploitation.
- The MRWF-GWO incorporates a novel position update strategy in which the alpha wolf's position is no longer influenced by beta and delta wolves with lower fitness. Likewise, the beta wolf's position update is independent of the delta wolf's low fitness value.
- The MRWF-GWO integrates the GWO with a median random walk wolf position, enhancing its ability to escape local optima. We propose utilizing both local and global search strategies to improve the search for optimal solutions.

3. Mathematical methods

This section discusses about the mathematical methods used in our research.

3.1. Grey wolf optimizer

GWO [60] is a nature-inspired, swarm-based optimization technique modeled on the social hierarchy and cooperative hunting strategies of grey wolves. Grey wolves live and hunt in packs, exhibiting a structured population hierarchy consisting of four levels. The alpha wolf holds the highest rank and leads the group followed by the beta and delta wolves, while omega wolves occupy the lowest rank. The alpha wolf directs the pack's movements and decisions, whereas the beta wolf acts as a subordinate to the alpha. The delta wolf ranks below the beta but supervises the omega wolves. As described by researchers in [60], the principal aspects of the pack's hunting behavior are outlined as follows:

Encircling the prey:

The alpha solution corresponds to the candidate with the highest fitness value (FV). The solution with the second-highest FV is termed the beta solution, while the third-highest FV is referred to as the delta solution. All remaining solutions are treated as omega solutions. These solutions are guided and improved based on the positions of the leading wolves. The encircling mechanism is defined as shown in Eq. (1)

$$Y^d(\tau+1) = Y_k^d(\tau) - A \times D \quad (1)$$

where $Y^d(\tau+1)$ and $Y_k^d(\tau)$ are the position vectors of each grey wolf and prey (k) respectively, d is the dimension of grey wolf at current iteration τ . D is the distance between prey and wolf which is obtained in Eq. (2).

$$D = |C \times Y_k^d(\tau) - Y^d(\tau)| \quad (2)$$

where A and C are coefficient vectors as given in Eqs. (3) and (4) respectively.

$$A = 2ar_1 - a \quad (3)$$

$$C = 2r_2 \quad (4)$$

Here, r_1 and r_2 are random numbers between $[0, 1]$. a is the convergence factor which is defined as in Eq. (5)

$$a = 2 - \frac{2\tau}{T} \quad (5)$$

where T is the maximum number of iterations.

Hunting:

To emulate the collective hunting behavior of a wolf pack, the positions of the top three wolves – alpha (α), beta (β), and delta (δ) – are recorded at each iteration as the best, second-best and third-best solutions respectively. The remaining wolves, Y_i^d , update their positions based on Y_α^d , Y_β^d and Y_δ^d according to the update equations in Eqs. (6)–(12).

$$D_\alpha = |C_1 \times Y_\alpha^d(\tau) - Y_i^d(\tau)| \quad (6)$$

$$D_\beta = |C_2 \times Y_\beta^d(\tau) - Y_i^d(\tau)| \quad (7)$$

$$D_\delta = |C_3 \times Y_\delta^d(\tau) - Y_i^d(\tau)| \quad (8)$$

$$Y_1^d(\tau+1) = Y_\alpha^d(\tau) - A_1 \times D_\alpha \quad (9)$$

$$Y_2^d(\tau+1) = Y_\beta^d(\tau) - A_2 \times D_\beta \quad (10)$$

$$Y_3^d(\tau+1) = Y_\delta^d(\tau) - A_3 \times D_\delta \quad (11)$$

$$Y_i^d(\tau+1) = \frac{Y_1^d(\tau+1) + Y_2^d(\tau+1) + Y_3^d(\tau+1)}{3} \quad (12)$$

where A_1 , A_2 , and A_3 are computed similarly as in Eq. (3). C_1 , C_2 and C_3 are computed similarly by Eq. (4).

Attacking the prey:

The grey wolves conclude the hunt by moving towards the prey until it ceases movement. The prey remains within a search-space-defined circle determined by the positions of the alpha (α), beta (β) and delta (δ) wolves. It is noted that when the random coefficient A falls within the range $[-1, 1]$, the wolves explore the search space.

3.2. Mathematical enhancement on GWO - MRWF-GWO

In Section 3 of GWO, it is explained how a grey wolf pack successfully hunts with the guidance of pack leaders α , β and δ . Gupta and Deep [64] introduced a novel random walk GWO to enhance exploration of the search space for continuous functions applied in benchmarks. Our research introduces GWO based on random wandering fitness. In this approach, leaders explore the search space through random walks and update their position and corresponding fitness function accordingly.

The proposed MRWF-GWO enhances performance by discarding the least effective wolves and re-initializing them at random positions. For discrete problems, wolves move within the binary hypercube with positions updated based on the alpha, beta, and delta wolves and mapped using an S-shaped transfer function, defined as follows:

$$Y^d(\tau+1)(S) = \begin{cases} \text{median if } \text{sigmoid}\left(\frac{Y_1^d + Y_2^d + Y_3^d}{3}\right) > \text{rand} \\ 0 \quad \text{Otherwise} \end{cases} \quad (13)$$

rand is a uniform distribution $U(0, 1)$. The parameters Y_1^d , Y_2^d , and Y_3^d are the position vectors of α , β , δ updated as given by Eqs. (14)–(16):

$$Y_1^d = Y_1^d \pm (UB - LB \times RW_\alpha + LB) \quad (14)$$

$$Y_2^d = Y_2^d \pm (UB - LB \times RW_\beta + LB) \quad (15)$$

$$Y_3^d = Y_3^d \pm (UB - LB \times RW_\delta + LB) \quad (16)$$

where Y_1^d, Y_2^d , and Y_3^d denote the position vectors of α , β and δ wolf respectively in the d-dimensional space as described in Eq. (1). UB and LB indicate the upper and lower bounds of the search space while RW represents the random walk step for α , β and δ grey wolf as in Eq. (17).

$$Y^d(\tau) \in [LB, UB], LB = \min(Y^d(\tau)), UB = \max(Y^d(\tau)) \quad (17)$$

The low-performing candidate solutions are repositioned randomly within the boundaries of the search space as described in Eq. (18).

$$Y^d(\tau + 1) = (UB - LB \times RW + LB) \quad (18)$$

RW_i can be defined for each α , β and δ wolf as:

$$RW_i = \begin{cases} \frac{1}{\pi(1 + Y_i^2)} & \text{if } SC \geq \text{rand} \\ 0 & \text{Otherwise} \end{cases} \quad (19)$$

Where $i = \alpha, \beta, \delta$ and Y_i is the position of i and SC is defined as in Eq. (20):

$$SC = 0.99 - \frac{(0.99 - 0.01)\tau^2}{T^2} \quad (20)$$

Scale factor (SC) regulates the scattering of the leading grey wolves within the search space. Initially, a large SC promotes extensive position changes for the alpha (α), beta (β), and delta (δ) wolves thereby facilitating exploration. As iterations progress, SC decreases, limiting position changes and allowing focused exploitation around the best solutions. Considering that feature selection is inherently multi-objective – seeking to maximize classification accuracy while minimizing the number of selected features – the fitness function f is formulated accordingly in Eq. (21).

$$f = \sigma_1 \xi + \sigma_2 \frac{|L - R|}{|R|} \quad (21)$$

where ξ is the classification quality of the classifier for the selected features, $|L|$ and $|R|$ are the total number of features, σ_1 and σ_2 are constants selected in the range $[0, 1]$ such that $\sigma_1 + \sigma_2 = 1$. The selection of features using the proposed MRWF-GWO technique will be discussed in the next subsection.

3.3. Proposed MRWF-GWO feature selection method

The flowchart for the MRWF-GWO algorithm designed for FS is shown in Fig. 1. The feature selection process for various datasets using the MRWF-GWO algorithm is as follows:

Step 1: Data preparation and initial parameters

- Initially the datasets contain features and labels.
- Initialize the MRWF-GWO parameters namely number of populations (N), number of runs (M), coefficient parameters α, A, C .

Step 2: Fitness Evaluation

- Compute the fitness values given in Eq. (21) for each individual grey wolf in the population. This function evaluates how well a given set of features perform.

Step 3: Optimization

- During the optimization process, select the optimal parameters for the MRWF-GWO algorithm from the fitness function in Eq. (21) guided by performance metrics.

- Compute the first, second and third optimal solutions based on the hunting behavior of MRWF-GWO and social hierarchy.
- Updates the values of a, A, C in Eqs. (3) and (4) and the wolf positions using proposed MRWF-GWO method.

Step 4: Convergence checks

Verify if the stopping condition defined by the maximum number of iterations is satisfied.

- If the stopping criteria are not met, continue the process.
- If the stopping criteria are met, continue to the final selection.

Step 5: Final selection

- Select the top individuals (features) from the optimized set of grey wolves to construct a feature subset for the classification task. This subset contains the features that best contribute to predictive performance based on the MRWF-GWO algorithm.

Step 6: Model training and evaluation

- The learning algorithms are trained using the chosen feature subset and subsequently used to predict outcomes on the test set.
- Evaluate the model accuracy using metrics such as F1 score, accuracy, fitness value, feature size and elapsed time for training.

Step 7: Model Explainability

- Apply XAI techniques like SHAP (SHapley Additive exPlanations) to interpret the model's predictions.

4. Performance metrics and experimental design

This section describes the benchmark evaluation metrics used in this study as well as the experimental datasets and hyper-parameter settings used for experimentation.

4.1. Evaluation indices

We systematically compute and analyze several key performance metrics: accuracy to evaluate the model's precision; average feature size to assess the complexity and efficiency of the model; elapsed time to gauge the computational speed; and standard deviation to understand the variability of our experimental results. Standard statistical metrics were used to validate the data. Each optimization algorithm was run 20 times ($M = 20$) under identical experimental conditions. The performance indicators for various datasets are described below:

Average Classification Accuracy (Acc): Accuracy quantifies the proportion of correctly classified instances in relation to the total number of samples. It is computed as shown in Eq. (22), where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (22)$$

μ_{Acc} is defined as average classification accuracy which is denoted by Eq. (23):

$$\mu_{Acc} = \frac{1}{M} \sum_{i=1}^M Acc^i \quad (23)$$

where Acc^i is the accuracy value in the i th run.

Average fitness: This metric evaluates the trade-off between feature subset reduction and the resulting improvement in classification accuracy. The fitness (μ_{Fit}) function is defined as in Eq. (24).

$$\mu_{Fit} = \frac{1}{M} \sum_{i=1}^M f^i \quad (24)$$

where f^i is the average fitness value in the i th run.

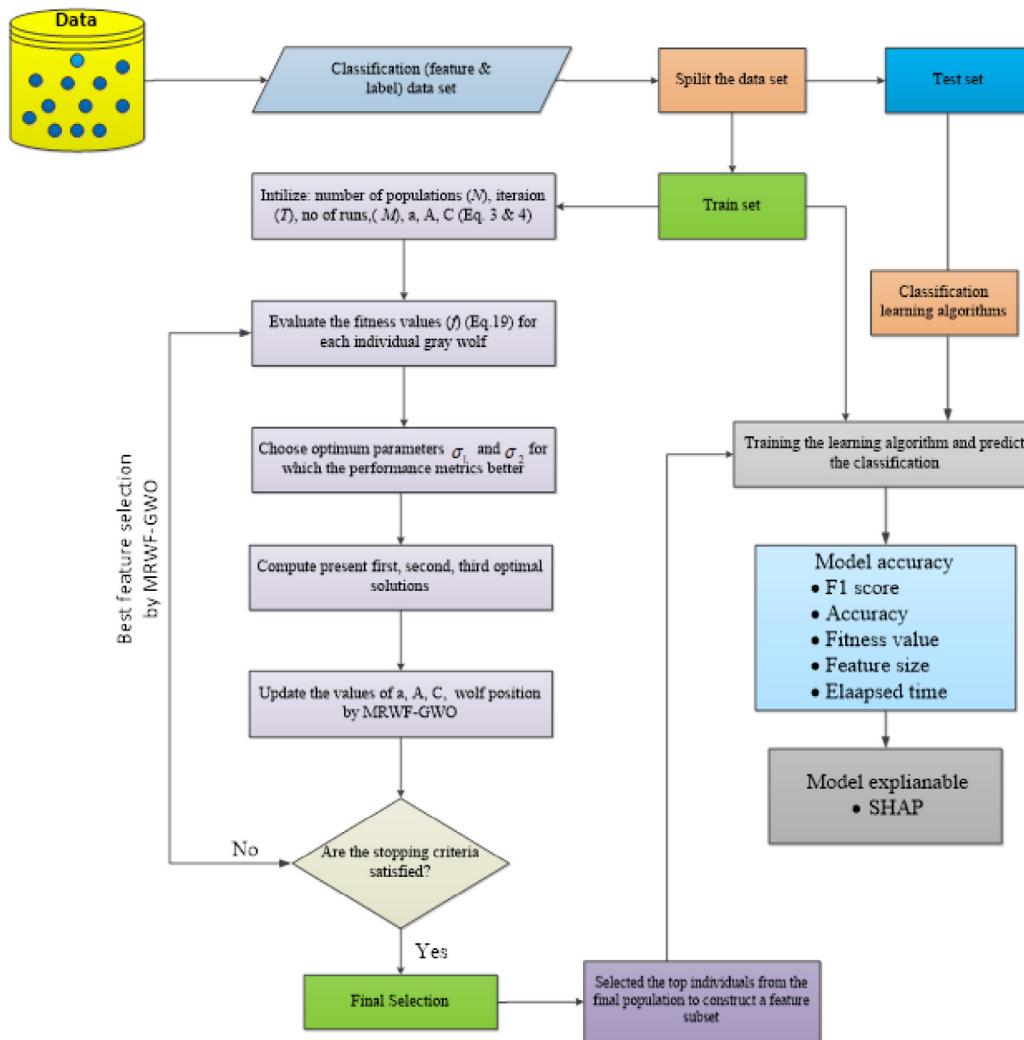


Fig. 1. Schematic diagram of feature selection by proposed MRWF-GWO algorithm for classification.

Average feature selection size: The average size of feature selection is denoted by μ_{FS} and defined in Eq. (25):

$$\mu_{FS} = \left\lceil \frac{1}{M} \sum_{i=1}^M f_i \right\rceil \quad (25)$$

where $\lceil \cdot \rceil$ is the greatest positive integer. The feature selection ratio is determined by the proportion of selected features f_i relative to the total features F in the dataset, as shown in Eq. (26).

$$Overall_{FS} = \frac{1}{M} \sum_{i=1}^M \frac{f^i}{F} \quad (26)$$

where f^i is the feature selection size in the i th run.

Average F1: The F1 score represents the harmonic balance between precision and recall, with values closer to 1 indicating higher accuracy. It is calculated as shown in Eq. (27).

$$F1score = \frac{2 \times precision \times recall}{precision + recall} \quad (27)$$

$precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$. The average F1 score is calculated, as in Eq. (28).

$$\mu_{F1} = \frac{1}{M} \sum_{i=1}^M F1score^i \quad (28)$$

where $F1score^i$ - F1 value in the i th run.

Standard deviation (SD): For metaheuristic optimization techniques, σ_x captures the variation in results across different runs (Eq. (29)).

$$\sigma_x = \sqrt{\frac{1}{M} \sum_{i=1}^M (f^i - \mu_{FS})^2} \quad (29)$$

All measurements including accuracy, best fitness, F1 score and size of selected features were computed with σ_x .

Average time period: The average computational time for each optimization method is measured using μ_{Time} as shown in Eq. (30).

$$\mu_{Time} = \frac{1}{M} \sum_{i=1}^M Time^i \quad (30)$$

where $Time^i$ is the time consumption in the i th run.

4.2. Experimental datasets

We have selected chronic diseases medical datasets to validate our proposed model. The proposed model requires environmental settings and parameter adjustments to ensure the best fitness function and better performance. This research operates on ten real-world datasets focusing on patient medical diagnoses. The datasets were obtained from repositories such as GitHub [65] and UCI. Table 2 summarizes their key characteristics and Table 3 lists the experimental parameters. Most datasets have binary labels while some contain 3–5 classes.

Table 2

Characteristics of the chronic medical datasets.

Dataset	Instances	Feature count	Classes
Glioma	50	4434	4
BCD	569	32	2
Diabetes	768	8	2
Hepatitis	155	23	2
Leukemia	72	7070	2
Kidney	400	19	2
Lung	203	3312	5
<i>CLL_SUB_111</i>	111	11,340	3
<i>GLI_85</i>	85	22,823	2
<i>Prostate_GE</i>	102	5966	2

- **Glioma:** Brain tumor dataset with four classes.
- **BCD (Breast Cancer):** Binary-class breast cancer dataset.
- **Diabetes:** Binary outcome (0 = negative, 1 = positive).
- **Hepatitis:** 155 samples, 23 features, binary classes (32 deceased (20.6%), 123 alive (79.4%)).
- **Leukemia:** Gene expression dataset with binary classes.
- **Kidney:** 400 samples for predicting chronic kidney disease, binary classes.
- **Lung:** 203 samples, five-class lung cancer dataset.
- **CLL_SUB_111:** Gene expression data obtained from oligonucleotide arrays with genetically and clinically distinct classes.
- **GLI_85:** Gene expression samples of diffuse infiltrating gliomas which are the common brain malignancy in adults, divided into two classes; includes high-grade glioblastoma multiform.
- **Prostate_GE:** Gene expression data from 102 samples of prostate cancer and normal controls, binary classes.

Glioma (50 samples, 4 classes) and Leukemia (72 samples), are relatively small and require careful handling to ensure statistical reliability. To address this we employed an 80% training and 20% testing split using stratified sampling to preserve class distributions across both subsets. Given the limited number of samples per class in some datasets (e.g., approximately 12 samples per class in Glioma), we repeated the splitting and evaluation process 20 times with different random seeds to reduce the effect of variability due to random sampling. The final results were reported as the mean and standard deviation of performance metrics. In addition to small-sample datasets, our study also included larger and high-dimensional datasets such as Diabetes (768 samples), BCD (569 samples), *CLL_SUB_111* (11,340 features) and *GLI_85* (22,823 features). These datasets were included specifically to evaluate the scalability and generalizability of the proposed MRWF-GWO algorithm across different data sizes and feature dimensionalities.

4.3. Parameters and environment setting

In both MRWF-GWO and BGWO algorithms, the value of σ_1, σ_2 (0.01, 0.99; 0.25, 0.75; 0.4, 0.6; 0.99, 0.01) parameters from objective function in the local opposition learning strategy as given by Eq. (21) affects the proposed algorithm baseline BGWO and are therefore determined experimentally.

Table 4 shows the optimal fitness values, accuracy, and F1 score for different σ_1, σ_2 tested across 20 independent runs. The best fitness value has obtained when $\sigma_1 = 0.01, \sigma_2 = 0.99$ on four test datasets (Diabetes, Hepatitis, *GLI_85*, *Prostate_GE*). The F1-score of Glioma, Leukemia, and *CLL_SUB_111* are 1, 0.98, 0.98, respectively in both $\sigma_1 = 0.01, \sigma_2 = 0.99$ and $\sigma_1 = 0.99, \sigma_2 = 0.01$. However, the F1-score performs much better when $\sigma_1 = 0.99, \sigma_2 = 0.01$ for all the chronic medical datasets. The highest average accuracy scores were observed

Table 3

Parameter setting of proposed feature selection MRWF-GWO and other MHA [Abbreviation: CR — Crossover rate, MR — Mutation rate, c1, c2: learning factor].

Algorithm	Control parameters	Value
MRWF-GWO	Search agent	10
	a	[0, 2]
	r1,r2	[0, 1]
	A	[-1, 1]
	C	[0, 2]
	T	100
	σ_1	0.99
	σ_2	0.01
	M	20
B-GWO	Search agent	10
	a	[0, 2]
	r1,r2	[0, 1]
	A	[-1, 1]
	C	[0, 2]
	T	100
	σ_1	0.99
	σ_2	0.01
	M	20
DE	Search agent	10
	Scaling factor (F)	0.8
	CR	0.8
FOA	Search agent	10
	Step Size	10
GA	Search agent,	10,
	CR,	0.8,
	MR	0.01
PSO	Search agent,	10,
	c1,	2,
	c2,	2,
	w	1
SCA	Search agent	10
	a	2
WOA	Search agent,	10,
	b	1
KNN	K	5

when $\sigma_1 = 0.99, \sigma_2 = 0.01$ for all datasets except *CLL_SUB_111*. However, both the fitness value and the F1 score were much better in this dataset. When $\sigma_1 = 0.99, \sigma_2 = 0.01$, enhancing the diversity of the initial population helps the MRWF-GWO algorithm avoid local optima and overcome search stagnation. Table 4 ensures that when $\sigma_1 = 0.99, \sigma_2 = 0.01$, the proposed model performs well with respect to average accuracy, average fitness and average F1 score in majority of datasets. Therefore, σ_1, σ_2 parameters are set to 0.99, 0.01 to guarantee the efficient search behavior of MRWF-GWO in high-dimensional feature space. We have fixed the parameters σ_1, σ_2 of our proposed fitness function in Eq. (21). Other parameters listed in Table 4 are also fixed. Since the proposed MRWF-GWO method is based on wrapper-based feature selection, the classification error rate is evaluated using the standard KNN algorithm with K = 5.

As observed from Table 4, MRWF-GWO demonstrates superior performance compared to the BGWO across all considered datasets and parameter combinations. MRWF-GWO achieves markedly higher classification accuracy and F1-scores accompanied by lower or comparable optimal fitness values. On the Glioma, Leukemia, Kidney and Lung datasets, MRWF-GWO consistently attains near-perfect F1-scores

Table 4

Comparative performance analysis of MRWF-GWO and BGWO algorithms in terms of the average fitness values (σ_{Fit}), accuracy (σ_{Acc}) and F1-score (σ_{F1}) for different values of σ_1 and σ_2 .

Data Set	$\sigma_1 = 0.01, \sigma_2 = 0.99$			$\sigma_1 = 0.25, \sigma_2 = 0.75$			$\sigma_1 = 0.40, \sigma_2 = 0.60$			$\sigma_1 = 0.99, \sigma_2 = 0.01$		
	σ_{Fit}	σ_{Acc}	σ_{F1}									
Glioma	0.1036	89.00	1.00	0.1099	88.50	0.98	0.1095	89.00	0.98	0.0248	97.50	1.00
BCD	0.0338	92.48	0.90	0.0438	92.48	0.90	0.0520	93.01	0.90	0.0429	95.75	0.94
Diabetes	0.1267	70.45	0.51	0.1677	70.42	0.50	0.1959	69.77	0.50	0.2217	77.79	0.64
Hepatitis	0.0583	66.77	0.53	0.1195	68.87	0.51	0.1635	68.71	0.50	0.2226	77.74	0.70
Leukemia	0.1170	91.79	0.98	0.1025	96.43	0.94	0.0880	97.14	0.95	0.0107	98.93	0.98
Kidney	0.0398	98.42	0.99	0.0335	99.81	1.00	0.0267	99.62	1.00	0.0011	99.94	1.00
Lung	0.0944	96.50	0.93	0.0811	99.13	0.98	0.0716	99.25	0.99	0.0051	99.50	0.99
CLL_SUB_111	0.1174	88.18	0.98	0.1465	86.82	0.97	0.1528	85.45	0.98	0.1121	86.82	0.98
GLI_85	0.0015	97.35	0.98	0.1133	97.65	0.98	0.1233	92.09	0.98	0.0118	98.82	0.99
Prostate_GE	0.0022	91.00	0.91	0.1116	92.75	0.92	0.1061	92.50	0.92	0.0248	97.50	0.97
BGWO												
Data Set	$\sigma_1 = 0.01, \sigma_2 = 0.99$			$\sigma_1 = 0.25, \sigma_2 = 0.75$			$\sigma_1 = 0.40, \sigma_2 = 0.60$			$\sigma_1 = 0.99, \sigma_2 = 0.01$		
	σ_{Fit}	σ_{Acc}	σ_{F1}									
Glioma	0.4132	79.77	0.80	0.4107	77.51	0.88	0.4053	79.45	0.96	0.3034	79.77	0.98
BCD	0.0641	93.84	0.81	0.0587	92.48	0.88	0.0527	94.81	0.90	0.0462	95.53	0.92
Diabetes	0.2165	71.25	0.51	0.2039	73.65	0.55	0.1865	74.17	0.59	0.1425	76.04	0.61
Hepatitis	0.1692	86.53	0.51	0.1452	87.22	0.58	0.1763	89.58	0.62	0.0470	95.32	0.65
Leukemia	0.2068	83.69	0.88	0.1870	89.30	0.90	0.1860	91.42	0.94	0.1441	94.28	0.96
Kidney	0.3157	95.61	0.95	0.2150	96.09	0.95	0.1201	96.51	0.97	0.0223	97.71	0.98
Lung	0.1822	84.51	0.95	0.1708	85.18	0.96	0.1653	87.50	0.97	0.0204	88.15	0.96
CLL_SUB_111	0.3053	78.04	0.94	0.2759	86.82	0.95	0.1680	85.45	0.95	0.1174	88.40	0.96
GLI_85	0.2101	87.53	0.89	0.2012	90.73	0.96	0.1705	91.23	0.97	0.0938	90.75	0.97
Prostate_GE	0.1256	70.10	0.88	0.1305	72.75	0.90	0.0955	76.54	0.91	0.0210	78.93	0.94

(0.98–1.00) and accuracy levels exceeding 97% whereas BGWO exhibits noticeably lower classification performance and higher variability in fitness outcomes. This improvement is primarily attributed to the medium-random wandering fitness mechanism incorporated into the classic BGWO framework. The adaptive wandering function dynamically regulates the exploration–exploitation trade-off, enabling search agents to explore a broader region of the solution space during early iterations and focus more effectively on local refinement in later stages. The fitness function update integrates controlled randomness which mitigates premature convergence and preserves population diversity, leading to more robust global search behavior. The integration of random fitness upgrading through a scaling factor and adaptive control parameters endows MRWF-GWO with enhanced convergence stability, better generalization capability and higher classification reliability across diverse datasets. In light of these consistent improvements, MRWF-GWO is exclusively considered for subsequent experimental investigations and comparative evaluations in the remaining numerical section and its corresponding subsections.

As metaheuristic methods are stochastic, each algorithm was run 20 times for consistency. The maximum number of iterations (T) is set to 100, the dimension to the total number of features (d), the total number of runs (M) to 20, and the number of search populations (N) to 10 for each iteration to ensure that the program runs smoothly. Experiments were conducted in MATLAB 2021 and Python 3.12.0 on an Intel(R) Xeon(R) W-2133 CPU @ 3.60 GHz with 64 GB RAM. The environment is set for the program, and parameters are fixed for further performance evaluation and experimental results.

5. Experimental results and discussion

Experimental results include a comprehensive analysis of performance metrics specifically investigating classification accuracy (in Section 5.1), average fitness value (in Section 5.2), convergence (in Section 5.3), average feature size (in Section 5.4), statistical analysis (in Section 5.5), SHAP explainability (in Section 5.6), elapsed time (in Section 5.7), and comparison with existing models (in Section 5.8).

5.1. Performance analysis of proposed MRWF-GWO and other MHAs

As shown in Table 5, MRWF-GWO attains the highest average classification accuracy on most datasets with the exception of Leukemia, where it ranks third after WOA and SCA. For the Kidney dataset, the average accuracy of 99.94% shows insignificant changes for the MRWF-GWO method compared to other metaheuristic algorithms (MHAs), but the standard deviation of 0.28 significantly changes compared to all other methods. For the BCD and *Prostate_GE* datasets, the minimum values for our proposed model are greater than those for the other methods. Table 5 shows that our proposed model is superior compared to other MHAs with respect to average classification accuracy and maximum values after the 20th iteration. Figs. 2(a–j) show box plots representing the interquartile range (IQR) and the accuracy of different MHAs. Outliers are represented as individual points. The length of the boxplots and outliers show that there can be some inconsistency in the results. Specifically, Figs. 2(a), 2(e), 2(f) and 2(g) indicate that MRWF-GWO algorithm often has a median accuracy above 90%. On the other hand, Fig. 2(c) showcases the varied performance of MRWF-GWO. From Fig. 2(d), it is observed that the MRWF-GWO algorithm appears to have a tighter IQR, indicating more consistent accuracy with higher median value compared to the other MHAs. Fig. 2(e) shows that all MHA achieved median accuracy rates over 90% with no low outliers, reflecting consistent high accuracy and robust performance on the leukemia dataset. Figs. 2(h) and 2(i) show a similar pattern but with slightly lower average accuracies. Fig. 2(j) shows that MRWF-GWO has a median accuracy of around 90% with one outlier for *Prostate_GE* data, suggesting that it is mostly reliable but may vary compared to the other MHAs displayed.

5.2. Fitness performance analysis of proposed MRWF-GWO and other MHAs

Table 6 shows average fitness and standard deviations for different algorithms across several datasets. MRWF-GWO stands out with the

Table 5
Average classification accuracy.

Methods	Glioma				BCD			
	μ_{Acc}	σ_{Acc}	MAX	MIN	μ_{Acc}	σ_{Acc}	MAX	MIN
MRWF-GWO	97.5	9.1766	100	80	95.75	1.6393	97.35	91.15
DE	92	8.3351	100	70	95.09	1.6612	97.35	92.04
FOA	90.5	8.8704	100	70	95.62	2.5757	99.12	90.27
GA	88	8.9443	100	70	94.42	1.5488	96.46	91.15
PSO	85	7.6089	100	70	94.56	1.9741	99.12	92.04
SCA	90	5.5012	100	70	95.53	1.8039	99.12	92.04
WOA	89	9.1191	100	70	95.49	1.3126	98.23	92.92
Methods	Diabetes				Hepatitis			
	μ_{Acc}	σ_{Acc}	MAX	MIN	μ_{Acc}	σ_{Acc}	MAX	MIN
MRWF-GWO	77.79	2.5576	83.01	72.9	77.74	7.5398	83.87	64.52
DE	76.63	2.312	81.05	73.2	73.87	6.6933	87.1	61.29
FOA	75.88	3.3656	81.05	68.63	77.1	5.0083	87.1	67.74
GA	77.55	2.7344	82.35	72.55	74.19	6.1026	87.1	64.52
PSO	74.12	2.005	77.12	71.24	75.48	4.7271	90.32	64.52
SCA	77.35	2.4936	81.7	73.2	77.1	7.0129	87.1	61.29
WOA	76.6	2.5587	81.05	70.59	71.94	6.7907	83.87	58.06
Methods	Leukemia				Kidney			
	μ_{Acc}	σ_{Acc}	MAX	MIN	μ_{Acc}	σ_{Acc}	MAX	MIN
MRWF-GWO	97.5	2.6168	100	78.57	99.94	0.283	100	98.73
DE	93.57	6.5136	100	78.57	97.78	1.9208	100	93.67
FOA	97.14	6.3041	100	78.57	99.18	1.2191	100	96.2
GA	92.14	7.9944	100	78.57	99.62	1.1392	100	94.94
PSO	97.5	5.8052	100	78.57	97.41	2.3761	100	92.41
SCA	97.86	3.5902	100	92.86	99.43	1.045	100	96.2
WOA	98.93	5.2973	100	92.86	99.75	0.5195	100	98.73
Methods	Lung				<i>CLL_SUB_111</i>			
	μ_{Acc}	σ_{Acc}	MAX	MIN	μ_{Acc}	σ_{Acc}	MAX	MIN
MRWF-GWO	99.5	2.4197	100	97.5	86.82	7.3589	95.45	77.27
DE	98.38	2.0318	100	95	71.82	7.606	90.91	54.55
FOA	99.13	1.4678	100	95	80.23	8.3782	95.45	63.64
GA	98.13	2.1267	100	92.5	75.23	7.7301	90.91	63.64
PSO	97.25	2.4197	100	92.5	79.77	9.4975	90.91	63.64
SCA	97.58	1.0259	100	95	86.05	4.7272	95.45	72.73
WOA	98.75	1.7206	100	95	82.95	6.235	86.36	63.64
Methods	GLI_85				Prostate_GE			
	μ_{Acc}	σ_{Acc}	MAX	MIN	μ_{Acc}	σ_{Acc}	MAX	MIN
MRWF-GWO	98.82	2.414	100	94.12	97.5	2.5649	100	75
DE	93.82	7.0057	100	82.35	91.25	7.2321	100	75
FOA	95.37	5.695	100	94.12	95.25	4.7226	100	85
hline GA	94.71	3.25	100	88.24	91.5	5.4047	100	80
PSO	87.65	8.0745	100	82.35	89.25	3.7258	95	85
SCA	98.24	3.3602	100	94.12	91.75	7.6563	100	95
WOA	97.94	2.8785	100	94.12	94.25	4.3755	100	85

lowest average fitness in the Glioma dataset, signifying better performance and reliability. It also consistently leads in the BCD and Diabetes datasets with low average fitness and competitive standard deviations. Meanwhile WOA algorithm has a much better average fitness value of 0.0107 for Leukemia dataset, with our proposed method being the third best, followed by FOA. Additionally, the standard deviation of 0.0025 is also lower compared to other MHAs, including MRWF-GWO. However, MRWF-GWO continues to outperform other MHAs in Hepatitis and Kidney datasets, with the lowest fitness values. The Kidney dataset is a highlight where the standard deviation of MRWF-GWO is 0.0029, showing very consistent performance. It leads to average fitness values for the Lung and *CLL_SUB_111* datasets. MRWF-GWO maintains the lowest fitness values of 0.0118 and 0.0248, along with a small standard deviation of 0.0239 and 0.0254. With the GLI_85 and Prostate_GE datasets respectively. It is demonstrated that MRWF-GWO shows strong performance across these medical datasets. The analysis from Table 6 shows that standard deviations vary across various MHAs, but the

average fitness values of our proposed method are comparatively better excluding the Leukemia dataset. Overall, MRWF-GWO consistently shows low average fitness across most datasets.

5.3. Analysis of convergence of proposed MRWF-GWO and other MHAs

Fig. 3 shows the convergence trends of average fitness values for MRWF-GWO and the comparative algorithms viz. PSO, DE, FOA, WOA, SCA and GA. As shown in Figs. 3(c), 3(f) and 3(g), MRWF-GWO demonstrates a noticeable performance advantage from the 20th iteration onward, which persists until convergence. In Fig. 3(d), the average fitness stabilizes around the 45th iteration, while Fig. 3(h) shows only a slight difference between MRWF-GWO and SCA. Figs. 3(i) and 3(j) further reveal a clear advantage from the 10th iteration onward. From Figs. 3(a)–3(j), MRWF-GWO shows a clear advantage from early to later iterations which is particularly evident from the 10th and 20th iterations onward in several datasets. In most cases, it achieves

faster and more stable convergence, effectively avoids local optima and outperforms the other algorithms across the iterative process.

5.4. Average feature size of proposed MRWF-GWO and other MHAs

Table 7 shows the average number of features selected by various algorithms across different datasets. The MRWF-GWO algorithm consistently chooses fewer features while maintaining or even improving classification accuracy. The number of features is considered as the greatest positive integer rather than a decimal number. For example, in the Glioma dataset, MRWF-GWO selected 28 features, equal to the smallest set chosen by the WOA algorithm highlighting its efficiency in feature reduction. In the BCD, Diabetes and Hepatitis datasets, the feature selection of MRWF-GWO is comparable to other algorithms. Selection of three or four features from the original feature set shows its strong feature reduction capabilities but insignificant difference for

other MHAs. However, **Table 7** shows that in the Leukemia dataset, MRWF-GWO achieves a 99% reduction from the original features, although the average accuracy (see **Table 5**) and average fitness (see **Table 6**) are better for WOA and SCA respectively. For the Kidney dataset, our proposed method reduces the features by approximately 95% from the original 23 features, although the number of features reduced by other methods such as GA and WOA is the same. The performance on the Lung, *CLL_SUB_111*, and *GLI_85* datasets further underscore the effectiveness of the MRWF-GWO algorithm by selecting significantly fewer features than competing algorithm (approximately 96% fewer), thereby highlighting the superiority of our proposed method. In the Prostate dataset, MRWF-GWO selected only 57 features, reducing approximately 99% from original feature set than other MHAs. The MRWF-GWO algorithm also obtained the lowest average number of features. Overall, **Table 7** demonstrates that MRWF-GWO is superior in reducing feature size across most datasets compared to all other MHAs.

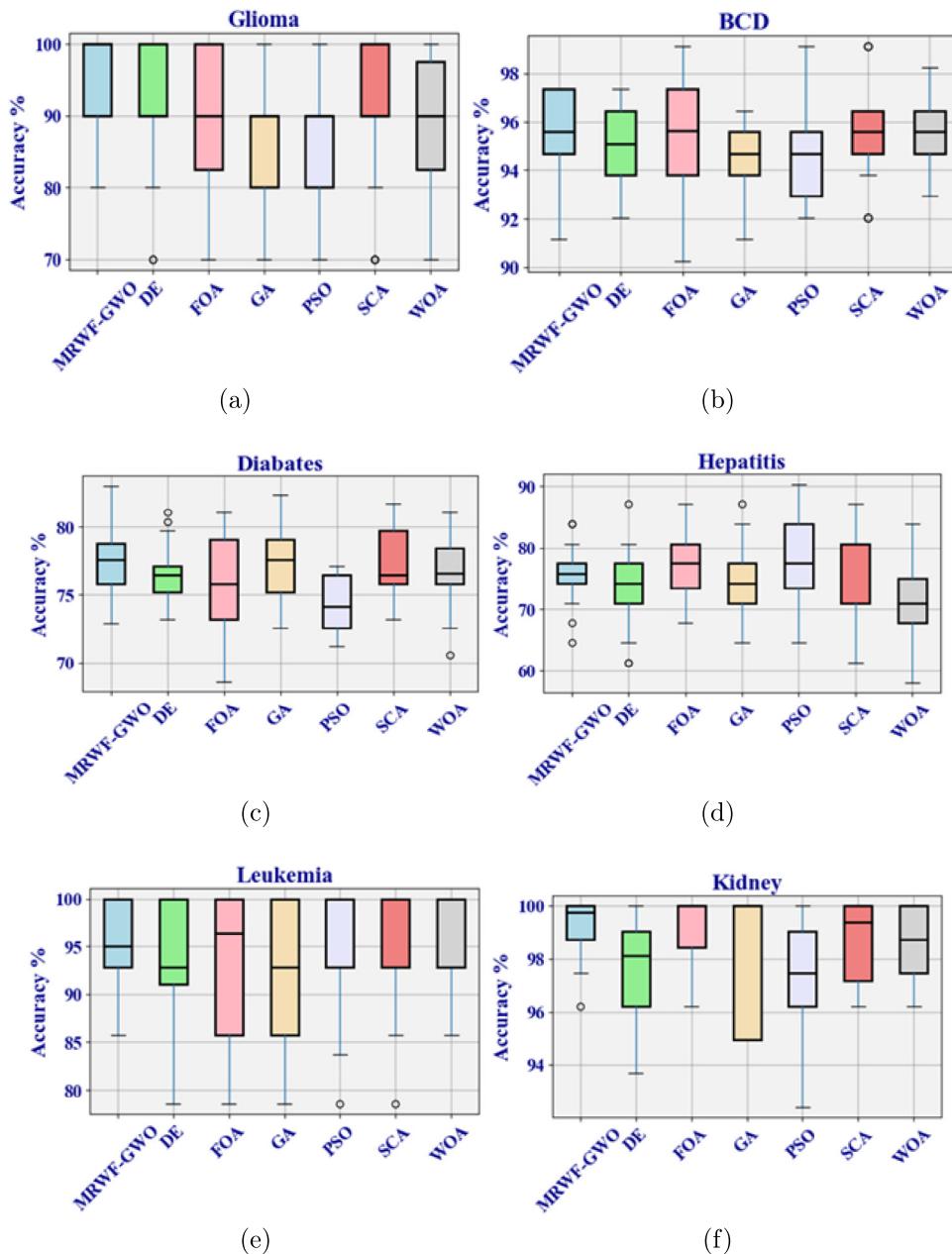


Fig. 2. Boxplots for classification accuracy of distinct algorithms - (a)-(j).

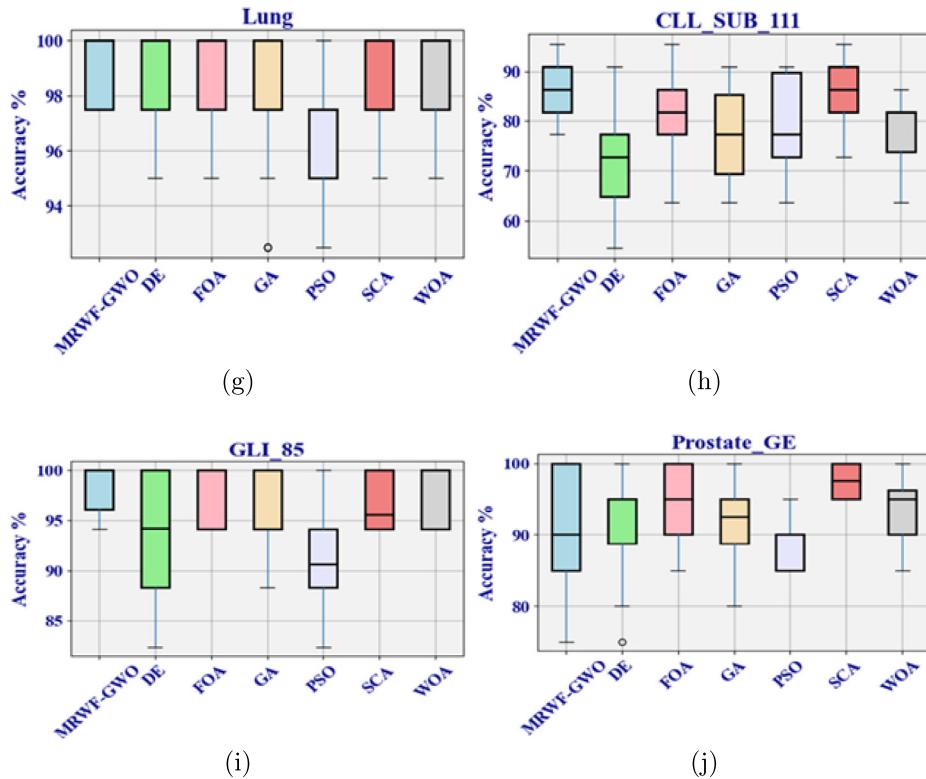


Fig. 2. (continued).

Table 6
Comparison of average fitness (μ_{Fit}) and standard deviation (σ_{Fit}).

Datasets	MRWF-GWO	DE	FOA	GA	PSO	SCA	WOA
Glioma	0.0248± 0.0908	0.0842± 0.0826	0.0948± 0.0878	0.1227± 0.0886	0.1531± 0.0753	0.1001± 0.0545	0.1090± 0.0902
BCD	0.0429± 0.0162	0.0521± 0.0164	0.0446± 0.0255	0.0562± 0.0153	0.0564± 0.0195	0.0452± 0.0179	0.0473± 0.0128
Diabetes	0.2217± 0.025	0.2369± 0.0227	0.2447± 0.0335	0.2273± 0.0268	0.2617± 0.0199	0.2296± 0.0245	0.2371± 0.0252
Hepatitis	0.2226± 0.0741	0.2630± 0.0659	0.2288± 0.049	0.2576± 0.06	0.2459± 0.0461	0.2285± 0.069	0.2796± 0.0668
Leukemia	0.0294± 0.0259	0.0687± 0.0644	0.0290± 0.0624	0.0820± 0.0791	0.0294± 0.0575	0.0107± 0.0025	0.0294± 0.0524
Kidney	0.0011± 0.0029	0.0242± 0.0193	0.0089± 0.0123	0.0043± 0.0115	0.0277± 0.0237	0.0062± 0.0105	0.0032± 0.0054
Lung	0.0051± 0.0239	0.0213± 0.0203	0.0094± 0.0145	0.0224± 0.0211	0.0318± 0.0239	0.0249± 0.0102	0.0134± 0.0169
CLL_SUB_111	0.1121± 0.0728	0.2852± 0.0751	0.1974± 0.083	0.2498± 0.0765	0.2051± 0.094	0.1198± 0.047	0.1692± 0.062
GLI_85	0.0118± 0.0239	0.0666± 0.0694	0.0124± 0.0239	0.0570± 0.0322	0.1271± 0.08	0.0190± 0.033	0.0208± 0.0283
Prostate_GE	0.0248± 0.0254	0.0918± 0.0717	0.0478± 0.0468	0.0883± 0.0535	0.1111± 0.0368	0.0829± 0.0758	0.0573± 0.0437

5.5. Wilcoxon rank-sum test

In the experimental study, seven benchmark datasets were employed to assess the performance of the proposed MRWF-GWO algorithm relative to DE, FOA, GA, PSO, SCA and WOA. Each algorithm was independently executed over 20 trial runs per dataset to ensure reliable performance evaluation. Statistical significance of the observed differences in accuracy was evaluated using the Wilcoxon rank-sum test (Mann–Whitney U test) [66]. For each dataset, the 20 accuracy values

obtained from MRWF-GWO were compared against the corresponding 20 values from each comparison algorithm to calculate the rank sums and the test statistics which were then used to derive the p-values. As a non-parametric method, the Wilcoxon rank-sum test does not assume normality in the underlying data distribution making it appropriate for non-Gaussian or ordinal datasets. By ranking the pooled data rather than relying on absolute values, the test provides a robust, distribution-free approach to determine whether the performance improvement of MRWF-GWO over other algorithms is statistically significant.

Table 7

Average retained feature size (μ_{FS}): The table presents the mean number of features selected by each algorithm on different datasets, with the proposed MRWF-GWO results emphasized.

Datasets	Original Feature Size	DE	FOA	GA	PSO	SCA	WOA-GWO	MRWF
Glioma	4434	1947	285	1518	1772	426	28	28
BCD	32	10	4	3	7	3	8	3
Diabetes	8	5	5	4	4	4	4	4
Hepatitis	23	8	4	4	6	3	3	4
Leukemia	7070	3547	483	2955	3275	67	3275	16
Kidney	23	6	2	1	5	1	2	1
Lung	3312	1636	235	1209	1429	1071	334	46
CLL_SUB_111	11,340	7070	1926	5134	5459	658	1866	538
GLI_85	22,823	12,156	1619	10,233	10,650	3321	953	331
Prostate GE	5966	3097	476	2476	2792	750	240	57

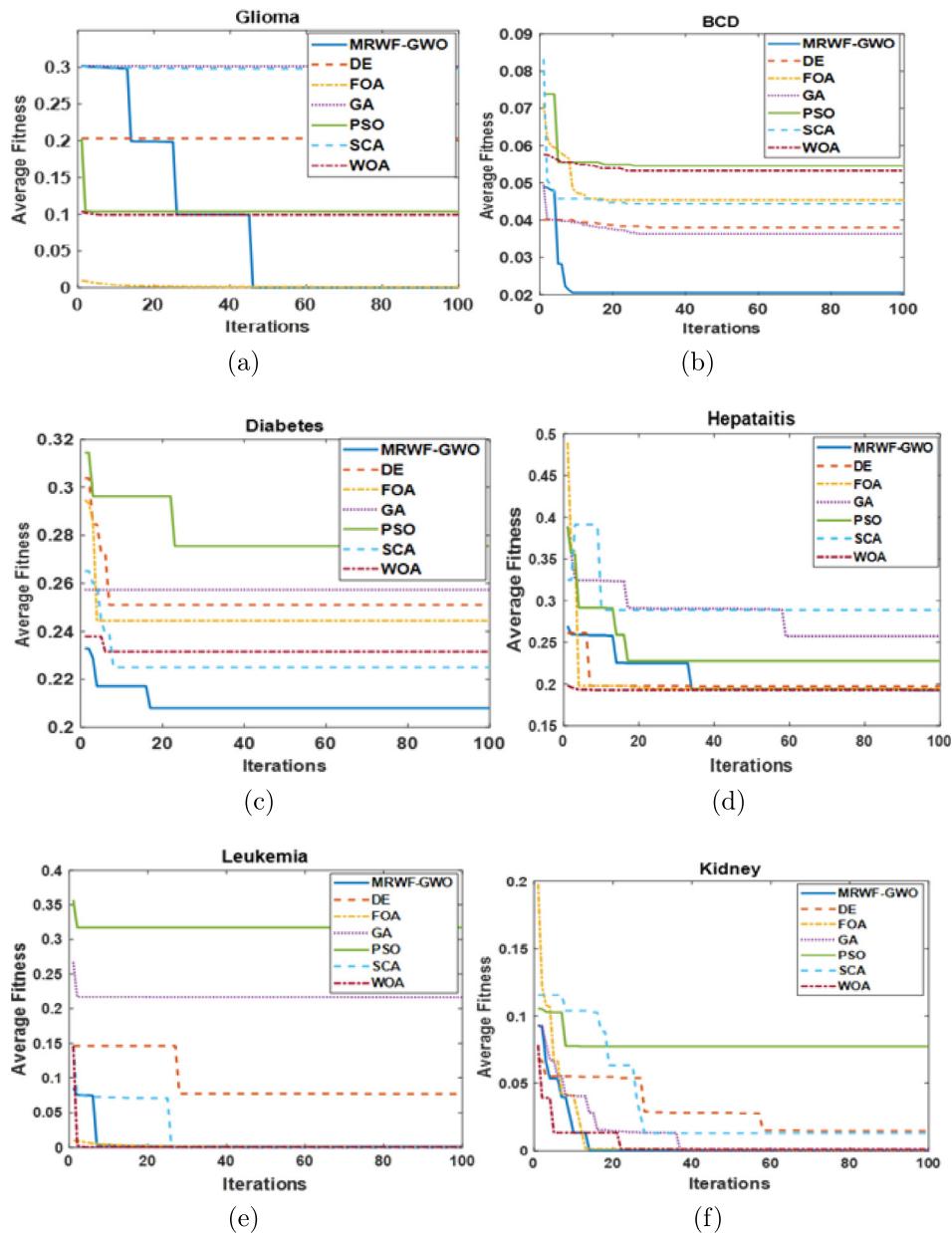


Fig. 3. Average fitness convergence plots of proposed MRWF-GWO and others MHAs across various medical datasets: (a)–(j).

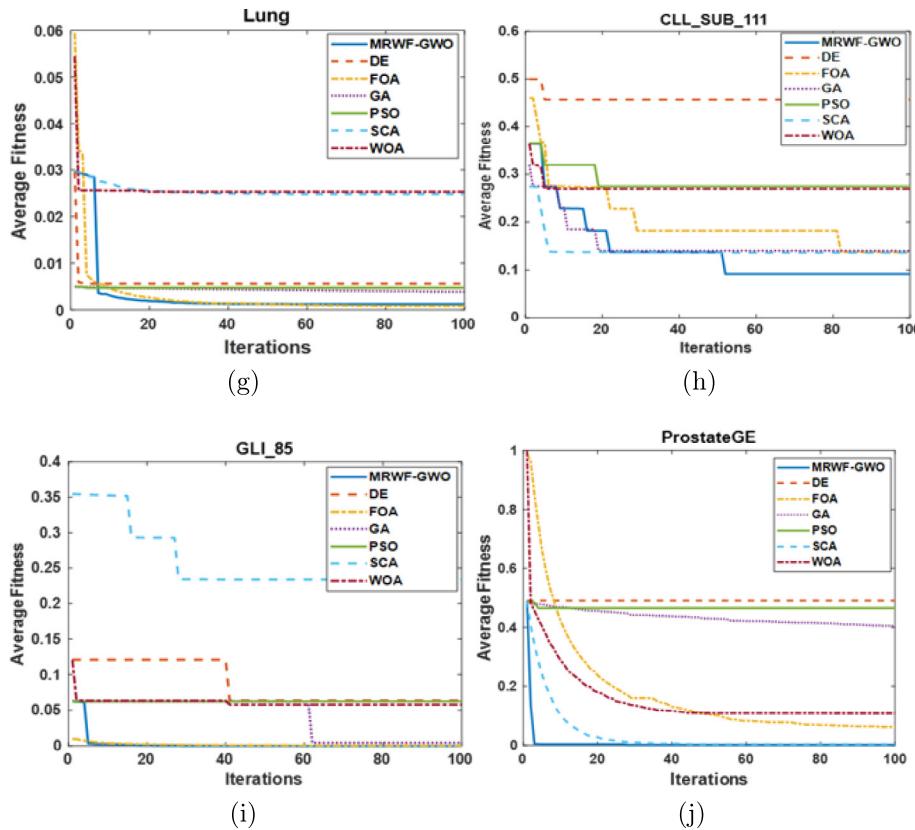


Fig. 3. (continued).

Table 8

Wilcoxon rank-sum test p-values comparing MRWF-GWO with other algorithms across ten benchmark datasets- since all p-values are much smaller than 0.05 (most being below 1×10^{-15}), their numerical comparison is unnecessary; they uniformly indicate overwhelming statistical significance in favor of MRWF-GWO.

Datasets	MRWF-GWO	DE	FOA	GA	PSO	SCA	WOA
Glioma	3.45E-19	4.13E-18	3.27E-18	2.26E-16	3.10E-17	6.65E-18	3.41E-18
h +	+	+	+	+	+	+	+
BCD	2.27E-20	3.87E-18	3.77E-18	4.12E-17	1.46E-13	2.74E-19	1.13E-16
h +	+	+	+	+	+	+	+
Diabetes	3.76E-20	3.22E-18	4.52E-18	4.25E-20	5.51E-17	4.87E-19	5.53E-19
h +	+	+	+	+	+	+	+
Hepatitis	3.88E-19	4.02E-18	4.07E-14	4.02E-18	6.85E-16	6.91E-18	4.14E-18
h +	+	+	+	+	+	+	+
Leukemia	2.76E-19	4.58E-11	5.76E-13	5.10E-16	4.36E-13	3.79E-16	1.28E-17
h +	+	+	+	+	+	+	+
Kidney	3.44E-19	8.00E-16	3.85E-18	3.22E-16	3.60E-18	3.87E-18	3.82E-18
h +	+	+	+	+	+	+	+
Lung	3.78E-20	3.90E-18	3.89E-16	3.98E-17	3.90E-18	5.73E-19	5.00E-18
h +	+	+	+	+	+	+	+
CLL_SUB_111	3.80E-20	6.61E-18	2.79E-17	3.70E-18	1.50E-17	7.52E-18	2.54E-17
h +	+	+	+	+	+	+	+
GLI_85	4.52E-19	4.67E-18	3.80E-18	6.67E-18	4.01E-18	3.55E-16	2.50E-17
h +	+	+	+	+	+	+	+
Prostate_GE	3.48E-20	4.20E-18	3.48E-19	3.70E-18	3.45E-18	3.64E-19	1.78E-17
h +	+	+	+	+	+	+	+

Each of the seven algorithms was executed independently 20 times on 10 benchmark datasets and the results were used for statistical comparison. A *p*-value below 0.05 indicates a significant difference while a value above this threshold suggests no statistically meaningful difference. MRWF-GWO is compared against DE, FOA, GA, PSO, SCA and WOA. Table 8 presents the corresponding *p*-values for these

comparisons. The results show that all *p*-values are far below 0.05 most being smaller than 10^{-15} which provides strong statistical evidence that the observed differences are unlikely to have occurred by chance. Therefore, Table 8 includes “+” symbols to indicate that MRWF-GWO performs significantly better than the compared algorithms. It should be noted that while all *p*-values are extremely small,

Table 9
Average elapsed time (μ_{Time}) in seconds.

Chronic Datasets	MRWF -GWO	DE	FOA	GA	PSO	SCA	WOA
Glioma	4.74	6.62	5.44	9.74	8.94	13.72	4.02
BCD	4.71	4.8	4.73	7.46	4.92	4.55	4.46
Diabetes	4.73	4.8	4.73	7.46	4.92	4.55	4.46
Hepatitis	4.22	4.38	4.27	6.69	4.24	4.08	3.71
Leukemia	4.33	10.54	6.69	15.07	14.87	7.97	14.87
Kidney	4.43	4.54	4.46	7.02	4.63	4.04	3.76
Lung	6.03	15.25	8.17	20.78	16.22	8.51	6.63
CLL_SUB_111	25.67	20.03	12.76	25.99	25.6	8.36	6.36
GLI_85	9.62	23.96	13.41	33.18	39.69	61.7	7.35
Prostate_GE	5.69	11.86	7.13	16.8	15.01	19.17	5.03

their numerical comparison carries no additional meaning as they all provide overwhelming evidence of MRWF-GWO's statistical superiority. These results confirm that MRWF-GWO achieves consistently superior performance compared to the other algorithms across all datasets. Overall, the statistically significant results across most comparisons affirm the superior performance of MRWF-GWO. Given that all p-values are already several orders of magnitude below the significance threshold, their precise numerical comparison has no additional interpretive value. They uniformly indicate the robustness and statistical superiority of MRWF-GWO.

5.6. Identifying top features through XAI using SHAP values

Beyond model transparency, the application of SHAP values substantially enhances diagnostic practices in real-world clinical settings. The alignment between SHAP-derived feature importance and established medical knowledge (e.g., glucose as a key predictor of diabetes, ascites as an indicator in hepatic conditions) strengthens clinician confidence in ML-driven insights. This interpretability allows clinicians to understand why a model arrives at a certain diagnosis, rather than treating it as a "black box". Such insights are particularly valuable in differential diagnosis, where nuanced interpretation of overlapping symptoms is essential. SHAP plots not only validate the algorithmic decisions with medically relevant justifications but also serve as a decision-support tool, facilitating data-driven yet clinically aligned interventions. Consequently, this approach supports greater trust, adoption and integration of AI models in healthcare diagnostics.

To develop an interpretable ML model for predicting chronic diseases, we employed an XAI approach using Shapley values to focus on features representing extreme values. This approach allowed us to assess the impact of specific features on disease classification. By leveraging Shapley values, it became feasible to identify the features crucial for accurate predictions. To elucidate the contribution of features within the MRWF-GWO model's prediction mechanism particularly through the KNN algorithm, SHAP values were computed for each feature. The importance of the features and observations are illustrated in Fig. 4.

Figs. 4a(i)–4j(i) display the SHAP plots of the top 10 features, showing their contributions to model predictions in descending order. The y-axis lists the features, x-axis represents the Shapley values and the color gradient indicates feature values from low to high.

Figs. 4a(i), 4d(i), 4g(i) and 4i(i) show that the features 'A33', 'ascites', 'A61' and 'GSM3921' contribute significantly to the model's predictions. Fig. 4c(i) portrays that the highest mean absolute SHAP value for diabetes was observed for Glucose, indicating it had the most significant impact, followed by 'Age', 'Pregnancies', 'BMI', 'DiabetesPedigree-Function', 'SkinThickness', 'BloodPressure', and 'Insulin' in that order. Similarly, all other datasets show their high and low mean absolute SHAP values indicating significant contributions corresponding to those datasets.

Figs. 4a(ii)–4j(ii) show SHAP summary plots highlighting positive and negative feature contributions. Features are ranked vertically by

importance. The vertical axis ranks features in order of decreasing importance while the horizontal axis reflects their influence on the model's predictive output. The color gradient indicates feature values with red denoting high significance and blue indicating low significance.

Fig. 4c(ii) demonstrates a scatter plot which explains the contribution of individual features to the output of a predictive model. SHAP values close to zero suggest a minimal impact while values further from zero indicate a stronger influence. Notably, 'Glucose' stands out with several points indicating a substantial positive impact on model predictions, hinting at its importance as a predictor. Similarly, 'BMI', 'Age' and 'BloodPressure' show wider spreads of SHAP values, reflecting their variable but significant roles. Similarly, Figs. 4d(ii)–4j(ii) illustrate the contributions of individual features to the model's predictions with the most important features prominently highlighted.

5.7. Time complexity analysis of MRWF-GWO

This section evaluates MRWF-GWO in terms of convergence speed. Although its execution time is relatively low, it is largely influenced by hardware and considered secondary. The time complexity includes various components: computing the median random walk fitness for leaders ($O(1)$) and updating the remaining wolves ($O(N)$). Thirdly, let N be the population size of the algorithm and D the number of features. The initialization step has a time complexity of $O(N \times D)$, as does the fitness evaluation. Updating the position of each grey wolf requires $O(N \times D \times D)$ operations. Consequently, the overall time complexity of the algorithm is $O(N \times D \times D \times T)$ where T is the number of iterations. This matches the time complexity of the standard GWO algorithm.

The algorithm exhibits a time complexity consistent with MRWF-GWO. Table 9 presents the average computational time (in seconds) for MRWF-GWO compared to other algorithms on all the datasets. WOA achieved the fastest execution time, closely followed by MRWF-GWO. However, MRWF-GWO has the fastest execution time for the Leukemia and Lung datasets.

5.8. Comparison with existing literature

Table 10 outlines a comparison of performance metrics across several studies focusing on different types of datasets including those used for the diagnosis of chronic diseases. Nevertheless, our proposed method demonstrates better performance than other existing methods, yielding an average accuracy of 92.89% across the ten chronic disease datasets considered in the study as shown in Table 10.

Hou et al. [35] developed BIFFOA with an S-shaped transfer function and tested it on four medical datasets such as Hepatitis, Leukemia and Lung cancer. The accuracy from these datasets was lower compared to those obtained with our MRWF-GWO. Although their model surpassed MRWF-GWO on the Leukemia dataset, it did not exceed the accuracy reported by Alweshah et al. [2]. Ji et al. [37] enhanced improved binary PSO (IBPSO) and incorporated the Lévy flight method,

testing it on five medical datasets. Within medical datasets, the BCD dataset showed an accuracy with an insignificant difference of 0.61 compared to the proposed MRWF-GWO. On the Lung dataset, MRWF-GWO achieved a significant improvement of 2.00 over the results reported by Ji et al. [37].

While Huang et al. [53] obtained high accuracies on Leukemia (97.48%), Lung (93.80%) and *Prostate_GE* (97.15%), MRWF-GWO outperformed their results on every dataset. The most notable improvement was in lung classification (+5.70%), while leukemia (+0.02%) and *Prostate_GE* (+0.35%) showed marginal gains. Singh & Singh [67] analyzed medical datasets for significant FS using a hybrid approach

that combines an ensemble of filter-based hybrid FS (EFHF) with KNN and SVM methods, applied to 22 datasets. The classification accuracies achieved by the proposed MRWF-GWO differed by 6% for BCD, 25% for Leukemia, 5% for Kidney, 18% for Lung and 18% for *Prostate_GE*. Nadimi-Shahraki et al. [41] introduced E-GWO for feature selection across nine distinct medical datasets. They reported 4% lower accuracy for Diabetes and Lung datasets than those achieved by MRWF-GWO. However, for the Hepatitis dataset, they achieved a 11% improvement in accuracy compared to the proposed method.

Kale & Yüzgeç [46] studied cosine algorithms (SCA) for feature selection across 10 medical datasets. They demonstrated high accuracy

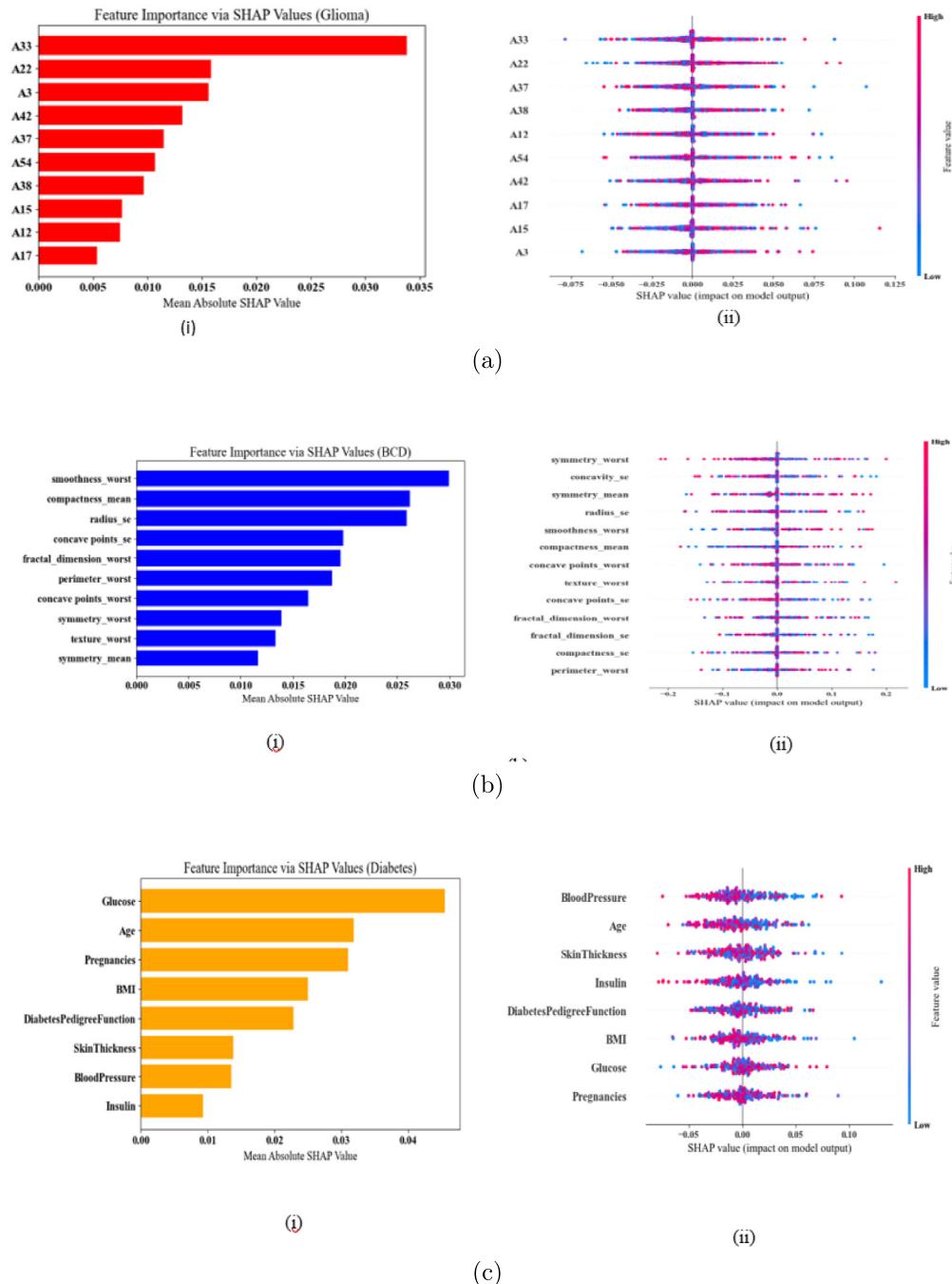


Fig. 4. The top 10 critical features as determined by SHAP values. Each subplot (a–j) shows: (i) the mean influence of each feature through horizontal bars, and (ii) the impact of the top 10 significant features on model predictions. Each dot corresponds to a subject; red = high feature value, blue = low, and the horizontal position indicates whether the feature increases or decreases the model's prediction.

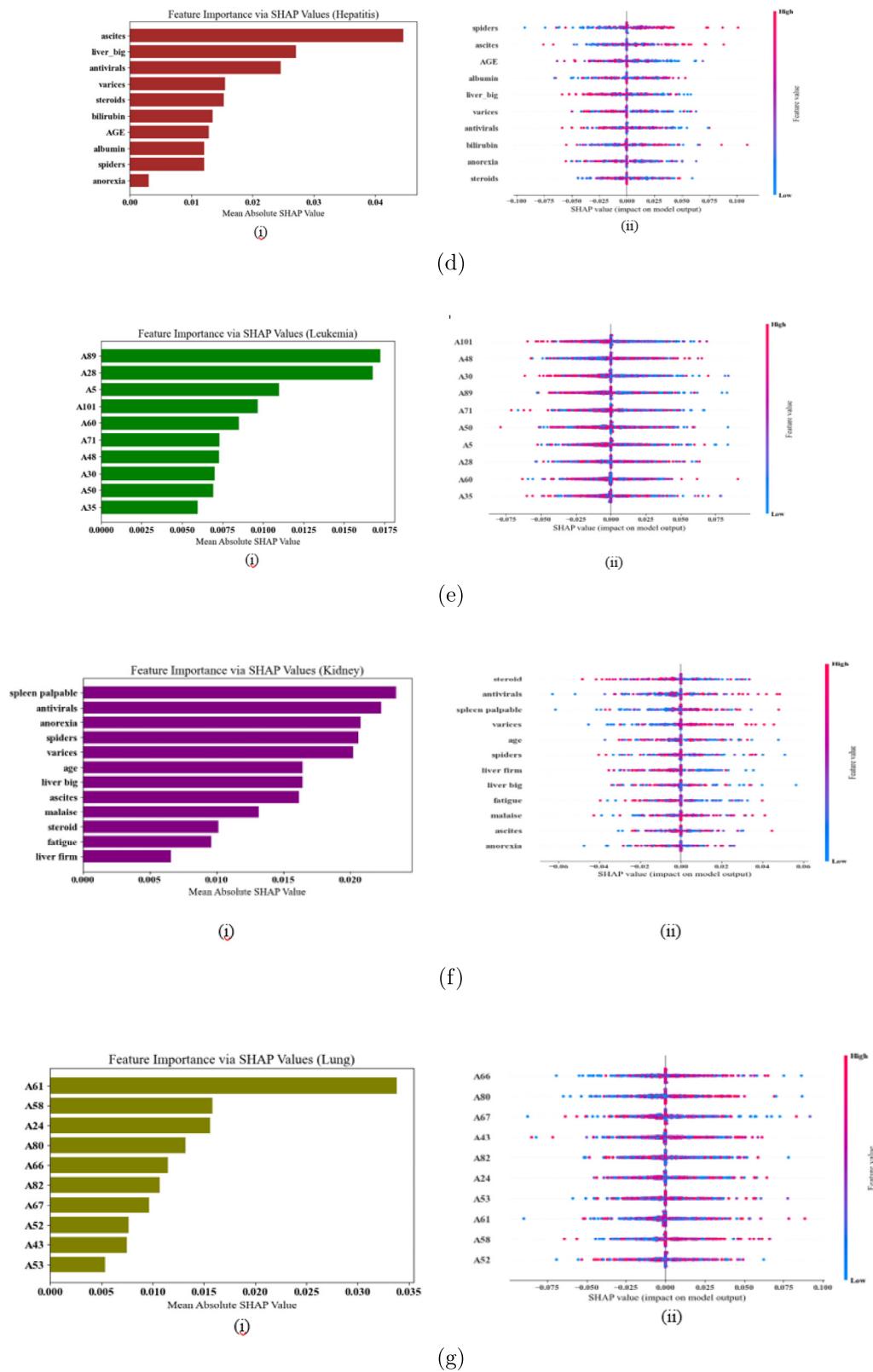


Fig. 4. (continued).

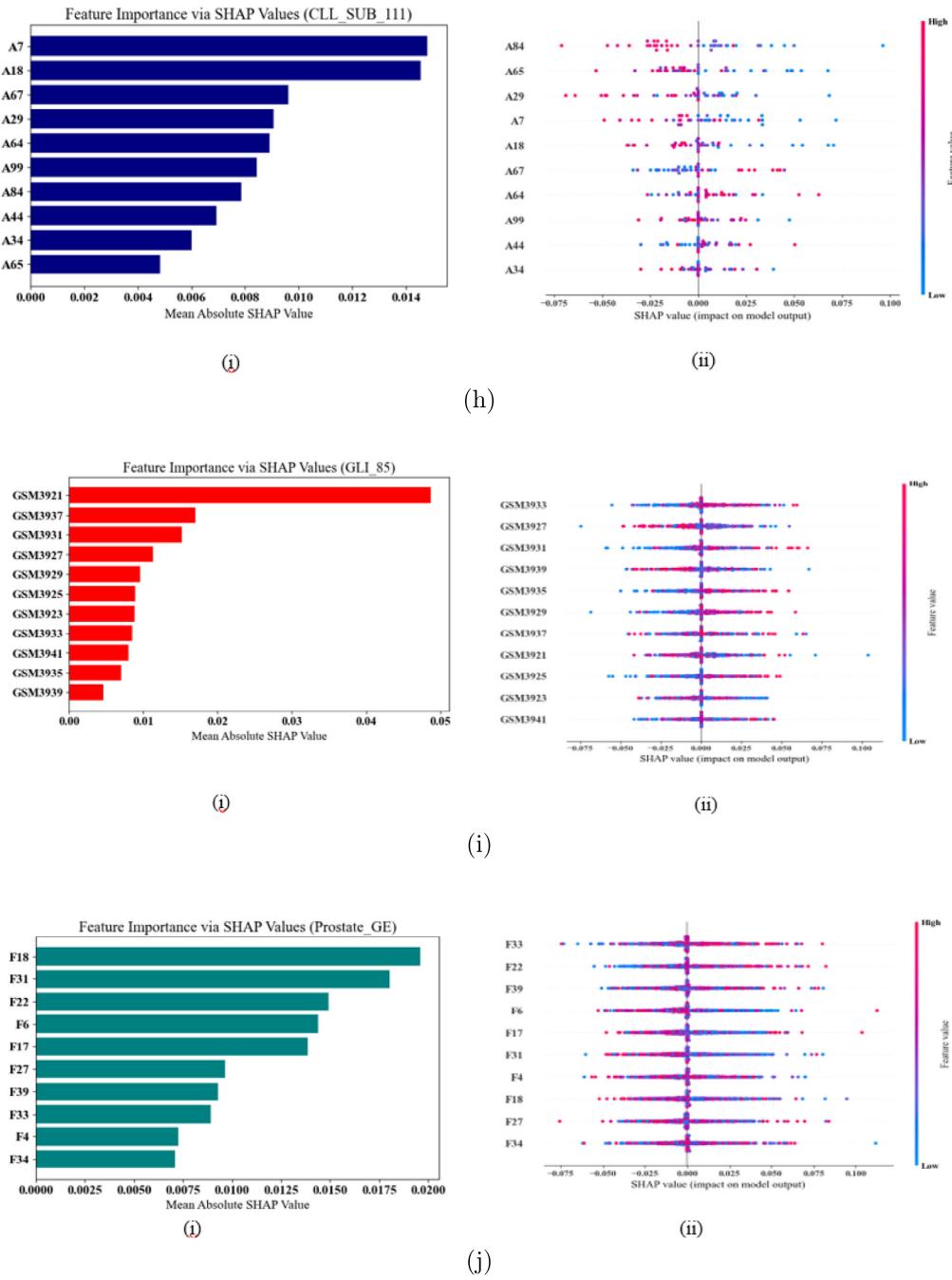


Fig. 4. (continued).

Table 10

Accuracy (%) comparison of proposed MRWF-GWO model with previous work. “–” indicates that the dataset was not considered in the corresponding study.

Chronic datasets	Hou et al. [35]	Ji et al. [37]	Huang et al. [53]	Singh and Singh [67]	Kale & Yuzgec [46]	Aziz et al. [68]	Alweshah et al. [69]	Houssein et al. [12]	Nadimi-Shahraki et al. [41]	Liu et al. [70]	MRWF-GWO
Glioma	–	–	–	–	–	87.05	–	–	–	–	97.5
BCD	–	95.14	–	90.12	93.5	–	94	98.54	–	–	95.75
Diabetes	–	–	–	–	–	–	74.56	–	75.93	–	77.79
Hepatitis	72.7	–	–	–	–	–	79.03	–	87.47	–	77.74
Leukemia	98.7	–	97.48	72.7	94.52	82.13	98.76	95.6	–	–	97.5
Kidney	–	–	–	95.1	–	–	–	–	–	–	99.94
Lung	96.9	97.5	93.8	81.2	–	79.66	–	–	95.52	96.57	99.5
CLL_SUB_111	–	–	–	–	–	–	–	–	–	–	86.82
GLI_85	–	–	–	–	–	–	–	–	–	91.81	98.82
Prostate_GE	–	–	97.15	79.75	–	83.82	60.1	87.16	–	93.09	97.5

rates for BCD and Leukemia, albeit lower than those achieved by MRWF-GWO. Aziz [68] investigated the nature-inspired cuckoo search (CS) algorithm, the Naive Bayes (NB) classifier and leave-one-out cross-validation (LOOCV) across six medical datasets covering Glioma, Leukemia, Lung and *Prostate_GE*. All of them recorded approximately 15% difference in accuracy than MRWF-GWO. Liu et al. [70] utilized a layer optimization method, incorporating SVM and KNN for FS across 6 medical datasets. While demonstrating strong performance in Lung Cancer, *GLI_85* and *Prostate_GE*, they were still lower compared to those achieved by MRWF-GWO. Houssein et al. [12] employed an enhanced Harris hawk's optimization with Lévy flight within 10 medical datasets. Among the algorithms tested, the accuracy for Leukemia and *Prostate_GE* were significantly lower compared to our model.

6. Conclusions and future scope

This study examined the MRWF-GWO method for feature selection in chronic medical datasets, benchmarking its performance against advanced feature selection techniques across several evaluation metrics. Experimental results demonstrated that MRWF-GWO outperformed competing methods in search capability, classification accuracy, feature subset size, stability, computational efficiency and convergence speed. MRWF-GWO yielded an average accuracy of 92.89% across the ten chronic disease datasets considered in the study. The algorithm considerably achieved a feature size reduction of approximately 95%–99% from original feature set. For Leukemia and Lung datasets, MRWF-GWO has the fastest execution time of 4.33s and 6.03s. Experimental results comparing optimal fitness values indicate that MRWF-GWO consistently surpasses the other MHAs in terms of faster convergence as well as accuracy and F1-score. This study also highlights the use of XAI to identify key features and analyze predictions using SHAP with a KNN classifier, revealing improvements in feature significance. While many MHAs perform well, overall performance depends on the classifier and no single method is optimal across all metrics.

In summary, the MRWF-GWO method offers significant benefits for feature selection. Nonetheless, it has limitations in reducing the feature dimension in specific datasets like Lung, *CLL_SUB_111*, *GLI_85* and *Prostate_GE*. Future work will explore incorporating approaches like federated and reinforcement learning to enhance position updates and global search. Filter-based methods such as mutual information (MI), chi-square and embedded methods like LASSO and decision tree-based selection are widely regarded as standard FS baselines particularly due to their low computational cost and statistical grounding. Future experiments will include both filter and embedded FS techniques to provide a more balanced perspective across FS paradigms and validate the generalizability and robustness of MRWF-GWO in broader systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge Sri Mata Amritanandamayi Devi (Amma), Chancellor, Amrita Vishwa Vidyapeetham, for her inspiration and for providing financial support for the Article Processing Charges (APC) of this publication.

Data availability

Datasets used in this study are available from the following repositories:

Scikit-feature benchmark datasets: Glioma, Kidney, Lung, *CLL_SUB_111*, *GLI_85*, *Prostate_GE*, Leukemia
<https://jundongli.github.io/scikit-feature/datasets.html>
 UCI Machine Learning Repository:
 Breast Cancer Wisconsin (Diagnostic) BCD <http://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>
 Hepatitis Dataset <http://archive.ics.uci.edu/dataset/46/hepatitis>
 Diabetes Dataset <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>.

References

- [1] W. Li, G.-G. Wang, A.H. Gandomi, A survey of learning-based intelligent optimization algorithms, *Arch. Comput. Methods Eng.* 28 (5) (2021) 3781–3799.
- [2] M. Alweshah, S. Alkhalaileh, D. Albashish, M. Mafarja, Q. Bsoul, O. Dorgham, A hybrid mine blast algorithm for feature selection problems, *Soft Comput.* 25 (2021) 517–534.
- [3] S. Sengan, G. Kamalam, J. Vellingiri, J. Gopal, P. Velayutham, V. Subramanyaswamy, et al., Medical information retrieval systems for e-health care records using fuzzy based machine learning model, *Microprocess. Microsyst.* (2020) 103344.
- [4] J.B. Raja, S.C. Pandian, PSO-fcm based data mining model to predict diabetic disease, *Comput. Methods Programs Biomed.* 196 (2020) 105659.
- [5] H. Bouzary, F.F. Chen, M. Shahin, Optimal composition of tasks in cloud manufacturing platform: a novel hybrid GWO-GA approach, *Procedia Manuf.* 34 (2019) 961–968.
- [6] L. Zhang, K. Mistry, S.C. Neoh, C.P. Lim, Intelligent facial emotion recognition using moth-firefly optimization, *Knowl.-Based Syst.* 111 (2016) 248–267.
- [7] F. Macedo, M.R. Oliveira, A. Pacheco, R. Valadas, Theoretical foundations of forward feature selection methods based on mutual information, *Neurocomputing* 325 (2019) 67–89.
- [8] A. Hashemi, M.B. Dowlatshahi, H. Nezamabadi-pour, A pareto-based ensemble of feature selection algorithms, *Expert Syst. Appl.* 180 (2021) 115130.
- [9] S. Acharya, L. Cui, Y. Pan, Multi-view feature selection for identifying gene markers: a diversified biological data driven approach, *BMC Bioinformatics* 21 (2020) 1–31.
- [10] F.A. Hashim, E.H. Houssein, K. Hussain, M.S. Mabrouk, W. Al-Atabany, A modified henry gas solubility optimization for solving motif discovery problem, *Neural Comput. Appl.* 32 (14) (2020) 10759–10771.
- [11] E.H. Houssein, I.E. Ibrahim, N. Neggaz, M. Hassaballah, Y.M. Wazery, An efficient ECG arrhythmia classification method based on manta ray foraging optimization, *Expert Syst. Appl.* 181 (2021) 115131.
- [12] E.H. Houssein, N. Neggaz, M.E. Hosney, W.M. Mohamed, M. Hassaballah, Enhanced harris hawks optimization with genetic operators for selection chemical descriptors and compounds activities, *Neural Comput. Appl.* 33 (2021) 13601–13618.
- [13] J. Cai, J. Luo, S. Wang, S. Yang, Feature selection in machine learning: A new perspective, *Neurocomputing* 300 (2018) 70–79.
- [14] R. Bellman, Dynamic programming, *Science* 153 (3731) (1966) 34–37.
- [15] Y. Zhang, S. Cheng, Y. Shi, D.-w. Gong, X. Zhao, Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm, *Expert Syst. Appl.* 137 (2019) 46–58.
- [16] Y. Zhang, D.-w. Gong, X.-z. Gao, T. Tian, X.-y. Sun, Binary differential evolution with self-learning for multi-objective feature selection, *Inform. Sci.* 507 (2020) 67–85.
- [17] X.-F. Song, Y. Zhang, Y.-N. Guo, X.-Y. Sun, Y.-L. Wang, Variable-size cooperative coevolutionary particle swarm optimization for feature selection on high-dimensional data, *IEEE Trans. Evol. Comput.* 24 (5) (2020) 882–895.
- [18] Y. Zhang, X.-f. Song, D.-w. Gong, A return-cost-based binary firefly algorithm for feature selection, *Inform. Sci.* 418 (2017) 561–574.
- [19] A. Alabert, A. Berti, R. Caballero, M. Ferrante, No-free-lunch theorems in the continuum, *Theoret. Comput. Sci.* 600 (2015) 98–106.
- [20] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary grey wolf optimization approaches for feature selection, *Neurocomputing* 172 (2016) 371–381.
- [21] H. Chantar, M. Mafarja, H. Alsawalqah, A.A. Heidari, I. Aljarah, H. Faris, Feature selection using binary grey wolf optimizer with elite-based crossover for arabic text classification, *Neural Comput. Appl.* 32 (2020) 12201–12220.
- [22] M. Abdel-Basset, D. El-Shahat, I. El-Henawy, V.H.C. De Albuquerque, S. Mirjalili, A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection, *Expert Syst. Appl.* 139 (2020) 112824.
- [23] Q. Al-Tashi, S.J.A. Kadir, H.M. Rais, S. Mirjalili, H. Alhussian, Binary optimization using hybrid grey wolf optimization for feature selection, *IEEE Access* 7 (2019) 39496–39508.

- [24] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Accelerating wrapper-based feature selection with K-nearest-neighbor, *Knowl.-Based Syst.* 83 (2015) 81–91.
- [25] N. Jiao, An efficient disease prediction framework based on optimized machine learning models for a smart healthcare application, *Multimedia Tools Appl.* 83 (17) (2024) 50825–50848.
- [26] M. Nssibi, G. Manita, O. Korbaa, Advances in nature-inspired metaheuristic optimization for feature selection problem: A comprehensive survey, *Comput. Sci. Rev.* 49 (2023) 100559.
- [27] S. Ansari, A.S. Alatrany, K.A. Alnajjar, T. Khater, S. Mahmoud, D. Al-Jumeily, A.J. Hussain, A survey of artificial intelligence approaches in blind source separation, *Neurocomputing* 561 (2023) 126895.
- [28] S. Ansari, K.A. Alnajjar, T. Khater, S. Mahmoud, A. Hussain, A robust hybrid neural network architecture for blind source separation of speech signals exploiting deep learning, *IEEE Access* (2023).
- [29] A. Holzinger, M. Dehmer, F. Emmert-Streib, R. Cucchiara, I. Augenstein, J. Del Ser, W. Samek, I. Jurisica, N. Díaz-Rodríguez, Information fusion as an integrative cross-cutting enabler to achieve robust, explainable, and trustworthy medical artificial intelligence, *Inf. Fusion* 79 (2022) 263–278.
- [30] E. Ephizibah, Cost effective approach on feature selection using genetic algorithms and fuzzy logic for diabetes diagnosis, *Int. J. Soft Comput.* 2 (2011).
- [31] A. Sahoo, S. Chandra, Multi-objective grey wolf optimizer for improved cervix lesion classification, *Appl. Soft Comput.* 52 (2017) 64–80.
- [32] L. Thara, R. Gunasundari, Adaptive feature selection method based on particle swarm optimization for gastric cancer prediction, in: Proc. Int. Conf. Comm. and Electronics Systems, ICICES, IEEE, 2017, pp. 808–813.
- [33] M.M. Mafarja, S. Mirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection, *Neurocomputing* 260 (2017) 302–312.
- [34] H. Faris, M.M. Mafarja, A.A. Heidari, I. Aljarah, A.-Z. Ala'm, S. Mirjalili, H. Fujita, An efficient binary salp swarm algorithm with crossover scheme for feature selection problems, *Knowl.-Based Syst.* 154 (2018) 43–67.
- [35] Y. Hou, J. Li, H. Yu, Z. Li, BIFFOA: A novel binary improved fruit fly algorithm for feature selection, *IEEE Access* 7 (2019) 81177–81194.
- [36] V. Enireddy, K. Gunda, N.L. Kalyani, K.B. Prakash, Nature inspired binary grey wolf optimizer for feature selection in the detection of neurodegenerative (PARKINSON) disease, *Int. J. 9* (3) (2020).
- [37] B. Ji, X. Lu, G. Sun, W. Zhang, J. Li, Y. Xiao, Bio-inspired feature selection: An improved binary particle swarm optimization approach, *IEEE Access* 8 (2020) 85989–86002.
- [38] R. Hans, H. Kaur, Opposition-based enhanced grey wolf optimization algorithm for feature selection in breast density classification, *Int J Mach Learn. Comput.* 10 (3) (2020) 458–464.
- [39] X. Zhang, Y. Xu, C. Yu, A.A. Heidari, S. Li, H. Chen, C. Li, Gaussian mutational chaotic fruit fly-built optimization and feature selection, *Expert Syst. Appl.* 141 (2020) 112976.
- [40] M. Wang, H. Chen, Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis, *Appl. Soft Comput.* 88 (2020) 105946.
- [41] M.H. Nadimi-Shahraki, H. Zamani, S. Mirjalili, Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study, *Comput. Biol. Med.* 148 (2022) 105858.
- [42] J. Jiang, Z. Zhao, Y. Liu, W. Li, H. Wang, DSGWO: An improved grey wolf optimizer with diversity enhanced strategy based on group-stage competition and balance mechanisms, *Knowl.-Based Syst.* 250 (2022) 109100.
- [43] S.K. Sahoo, A.K. Saha, A hybrid moth flame optimization algorithm for global optimization, *J. Bionic Eng.* 19 (5) (2022) 1522–1543.
- [44] J. Wang, D. Lin, Y. Zhang, S. Huang, An adaptively balanced grey wolf optimization algorithm for feature selection on high-dimensional classification, *Eng. Appl. Artif. Intell.* 114 (2022) 105088.
- [45] M.H. Nadimi-Shahraki, S. Taghian, S. Mirjalili, An improved grey wolf optimizer for solving engineering problems, *Expert Syst. Appl.* 166 (2021) 113917.
- [46] G.A. Kale, U. Yüzgeç, Advanced strategies on update mechanism of Sine cosine optimization algorithm for feature selection in classification problems, *Eng. Appl. Artif. Intell.* 107 (2022) 104506.
- [47] R.R. Rajammal, S. Mirjalili, G. Ekambaram, N. Palanisamy, Binary grey wolf optimizer with mutation and adaptive k-nearest neighbour for feature selection in parkinson's disease diagnosis, *Knowl.-Based Syst.* 246 (2022) 108701.
- [48] K. Deep, et al., A random walk grey wolf optimizer based on dispersion factor for feature selection on chronic disease prediction, *Expert Syst. Appl.* 206 (2022) 117864.
- [49] Z. Li, A local opposition-learning golden-sine grey wolf optimization algorithm for feature selection in data classification, *Appl. Soft Comput.* 142 (2023) 110319.
- [50] Y. Ou, P. Yin, L. Mo, An improved grey wolf optimizer and its application in robot path planning, *Biomimetics* 8 (1) (2023) 84.
- [51] H. Askr, M. Abdel-Salam, A.E. Hassani, Copula entropy-based golden jackal optimization algorithm for high-dimensional feature selection problems, *Expert Syst. Appl.* 238 (2024) 121582.
- [52] X. Sun, S. Guo, S. Liu, J. Guo, B. Du, A correlation-redundancy guided evolutionary algorithm and its application to high-dimensional feature selection in classification, *Neural Process. Lett.* 56 (2) (2024) 86.
- [53] J. Huang, X. Deng, L. Hu, Enhanced grey wolf optimizer with hybrid strategies for efficient feature selection in high-dimensional data, *Inform. Sci.* (2025) 121958.
- [54] M. Ihsan, F. Din, K.Z. Zamli, Y.Y. Ghadi, T.J. Alahmadi, N. Innab, A modified grey wolf optimizer with multi-solution crossover integration algorithm for feature selection, *J. Ambient. Intell. Humaniz. Comput.* 16 (1) (2025) 329–345.
- [55] M. Raveenthini, R. Lavanya, R. Benitez, Interpretable diagnostic system for multiocular diseases based on hybrid meta-heuristic feature selection, *Comput. Biol. Med.* 184 (2025) 109486.
- [56] M. Maazalahi, S. Hosseini, Machine learning and metaheuristic optimization algorithms for feature selection and botnet attack detection, *Knowl. Inf. Syst.* (2025) 1–49.
- [57] Z. Sadeghian, E. Akbari, H. Nematzadeh, H. Motameni, A review of feature selection methods based on meta-heuristic algorithms, *J. Exp. Theor. Artif. Intell.* 37 (1) (2025) 1–51.
- [58] J.Y. Khaseeb, A. Keshk, A. Youssef, Improved binary grey wolf optimization approaches for feature selection optimization, *Appl. Sci.* 15 (2) (2025) 489.
- [59] A.A. Ewees, M.M. Alshahrani, A.M. Alharthi, M.A. Gaheen, Optimizing feature selection and remote sensing classification with an enhanced machine learning method, *J. Supercomput.* 81 (2) (2025) 370.
- [60] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Adv. Eng. Softw.* 69 (2014) 46–61.
- [61] A.K. Dey, G.P. Gupta, S.P. Sahu, A metaheuristic-based ensemble feature selection framework for cyber threat detection in IoT-enabled networks, *Decis. Anal. J.* 7 (2023) 100206.
- [62] N. Sharma, L. Malviya, A. Jadhav, P. Lalwani, A hybrid deep neural net learning model for predicting coronary heart disease using randomized search cross-validation optimization, *Decis. Anal. J.* 9 (2023) 100331.
- [63] S. Alam, X. Zhao, I.K. Niazi, M.S. Ayub, M.A. Khan, A comparative analysis of global optimization algorithms for surface electromyographic signal onset detection, *Decis. Anal. J.* 8 (2023) 100294.
- [64] S. Gupta, K. Deep, A novel random walk grey wolf optimizer, *Swarm Evol. Comput.* 44 (2019) 101–112.
- [65] J. Li, Scikit-feature datasets, 2025, URL <https://jundongli.github.io/scikit-feature/datasets.html>. (Accessed 15 May 2024).
- [66] F. Wilcoxon, Individual comparisons by ranking methods, in: *Breakthroughs in Statistics: Methodology and Distribution*, Springer, 1992, pp. 196–202.
- [67] N. Singh, P. Singh, A hybrid ensemble-filter wrapper feature selection approach for medical data classification, *Chemometr. Intell. Lab. Syst.* 217 (2021) 104396.
- [68] R.M. Aziz, Cuckoo search-based optimization for cancer classification: A new hybrid approach, *J. Comput. Biol.* 29 (6) (2022) 565–584.
- [69] M. Alweshah, S. Alkhalaileh, M.A. Al-Betar, A.A. Bakar, Coronavirus herd immunity optimizer with greedy crossover for feature selection in medical diagnosis, *Knowl.-Based Syst.* 235 (2022) 107629.
- [70] J. Liu, D. Li, W. Shan, S. Liu, A feature selection method based on multiple feature subsets extraction and result fusion for improving classification performance, *Appl. Soft Comput.* 150 (2024) 111018.