Journal Pre-proof

AltWOA: Altruistic Whale Optimization Algorithm for feature selection on microarray datasets

Rohit Kundu, Soham Chattopadhyay, Erik Cuevas, Ram Sarkar

PII: S0010-4825(22)00141-X

DOI: https://doi.org/10.1016/j.compbiomed.2022.105349

Reference: CBM 105349

To appear in: *Computers in Biology and Medicine*

Received Date: 2 October 2021

Revised Date: 22 February 2022

Accepted Date: 22 February 2022

Please cite this article as: R. Kundu, S. Chattopadhyay, E. Cuevas, R. Sarkar, AltWOA: Altruistic Whale Optimization Algorithm for feature selection on microarray datasets, *Computers in Biology and Medicine* (2022), doi: https://doi.org/10.1016/j.compbiomed.2022.105349.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier Ltd.

Journal Pre-proof

AltWOA: Altruistic Whale Optimization Algorithm for Feature Selection on Microarray Datasets

Rohit Kundu

Department of Electrical Engineering, Jadavpur University
188, Raja S.C. Mullick Road, Kolkata-700032, West Bengal, INDIA
Email: rohitkunduju@gmail.com

Soham Chattopadhyay

Department of Electrical Engineering, Jadavpur University
188, Raja S.C. Mullick Road, Kolkata-700032, West Bengal, INDIA
Email: chattopadhyaysoham99@gmail.com

Erik Cuevas*

Departamento de Electrónica, Universidad de Guadalajara
CUCEI, Av. Revolución 1500, Guadalajara, Jal, MEXICO
Email: erik.cuevas@cucei.udg.mx

Ram Sarkar

Department of Computer Science and Engineering, Jadavpur University
188, Raja S.C. Mullick Road, Kolkata-700032, West Bengal, INDIA
Email: ramjucse@gmail.com

*Corresponding author: Erik Cuevas
*Corresponding author email: erik.cuevas@cucei.udg.mx

Journal Pre-proof

Journal Pre-proof

AltWOA: Altruistic Whale Optimization Algorithm for Feature Selection on Microarray Datasets

Rohit Kundu [a], Soham Chattopadhyay [b], Erik Cuevas[c,*], Ram Sarkar [d]

[a] Department of Electrical Engineering, Jadavpur University, Kolkata-700032, INDIA; rohitkunduju@gmail.com
[b] Department of Electrical Engineering, Jadavpur University, Kolkata-700032, INDIA; chattopadhyaysoham99@gmail.com
[c] Departamento de Electrónica, Universidad de Guadalajara, CUCEI, Av. Revolución 1500, Guadalajara, Jal, MEXICO; erik.cuevas@cucei.udg.mx
[d] Department of Computer Science & Engineering, Jadavpur University, Kolkata-700032, INDIA; ramjucse@gmail.com

Abstract

The data-driven modern era has enabled the collection of large amounts of biomedical and clinical data. DNA microarray gene expression datasets have mainly gained significant attention to the research community owing to their ability to identify diseases through the “bio-markers" or specific alterations in the gene sequence that represent that particular disease (for example, different types of cancer). However, gene expression datasets are very high-dimensional, while only a few of those are “bio-markers". Meta-heuristic-based feature selection effectively filters out only the relevant genes from a large set of attributes efficiently to reduce data storage and computation requirements. To this end, in this paper, we propose an Altruistic Whale Optimization Algorithm (AltWOA) for the feature selection problem in high-dimensional microarray data. AltWOA is an improvement on the basic Whale Optimization Algorithm. We embed the concept of altruism in the whale population to help efficient propagation of candidate solutions that can reach the global optima over the iterations. Evaluation of the proposed method on eight high dimensional microarray datasets reveals the superiority of AltWOA compared to popular and classical techniques in the literature on the same datasets both in terms of accuracy and the final number of features selected. The relevant codes for the proposed approach are available publicly at https://github.com/Rohit-Kundu/AltWOA.

Keywords: Feature Selection, Evolutionary meta-heuristic, Altruism, Cancer Detection, Gene Expression, Microarray Data

1. Introduction

The technological advancements in the modern world have led to revolutionary changes in data acquisition techniques. This, in turn, has also led to the presence of redundant or misleading data, which sometimes degrade the performance of a learning model. Further, along with the “curse of dimensionality" [1], such a large number of attributes in a dataset tend to overfit machine learning models. One such domain is the analysis of high-dimensional microarray datasets of gene expression, which helps identify the genes responsible for cancer. Cancer is caused by the irregular mutation of genes (in most cases, the TP53 gene is missing or damaged). This can lead to uncontrolled cell replication forming tumors that might spread to other organs (malignancy) and impair their regular functioning.

Gene expression datasets generally contain information of more genes than there are tissue samples available. Hence, for the pathologists to extract information on the genetics of the disease (like cancer), it is required to sift out only the relevant genes from the data. However, only a few of these genes from a dataset consisting of thousands of genes are relevant to the cancer diagnosis. Feature selection is a concept in machine learning that aims to eliminate redundancy from the feature space by selecting an optimal subset of features. However, exhaustive search for an optimal feature set requires ‘$(2^{No.\ of\ features})$’ evaluations, making it an NP-hard problem. Thus intelligent approaches have been devised for selecting the optimal subset while being computationally efficient.

Feature selection techniques can be broadly categorized into the following classes:

(a) *Filter Methods*: These methods use the intrinsic properties of the data to select the optimal feature subset without using any learning algorithm in its crux. These methods are computationally much more efficient than the other two categories. Examples of such techniques are the method of Pasi Luukka [2], and the Relief feature selection algorithm [3].

(b) *Wrapper Methods*: Such methods use a learning algorithm in its core to evaluate possible feature subsets and evolve to an optimal solution through multiple iterations. Examples of such methods are Whale Optimization Algorithm (WOA) [4], and the Binary Bat Algorithm [5].

(c) *Embedded Methods*: These methods perform the feature selection while model training, i.e., the feature selection and the classification, are performed simultaneously. Examples of such methods are CS-SVM [6], and the process by Guyon et al. [7].

Amongst the three feature selection techniques mentioned

*Corresponding author

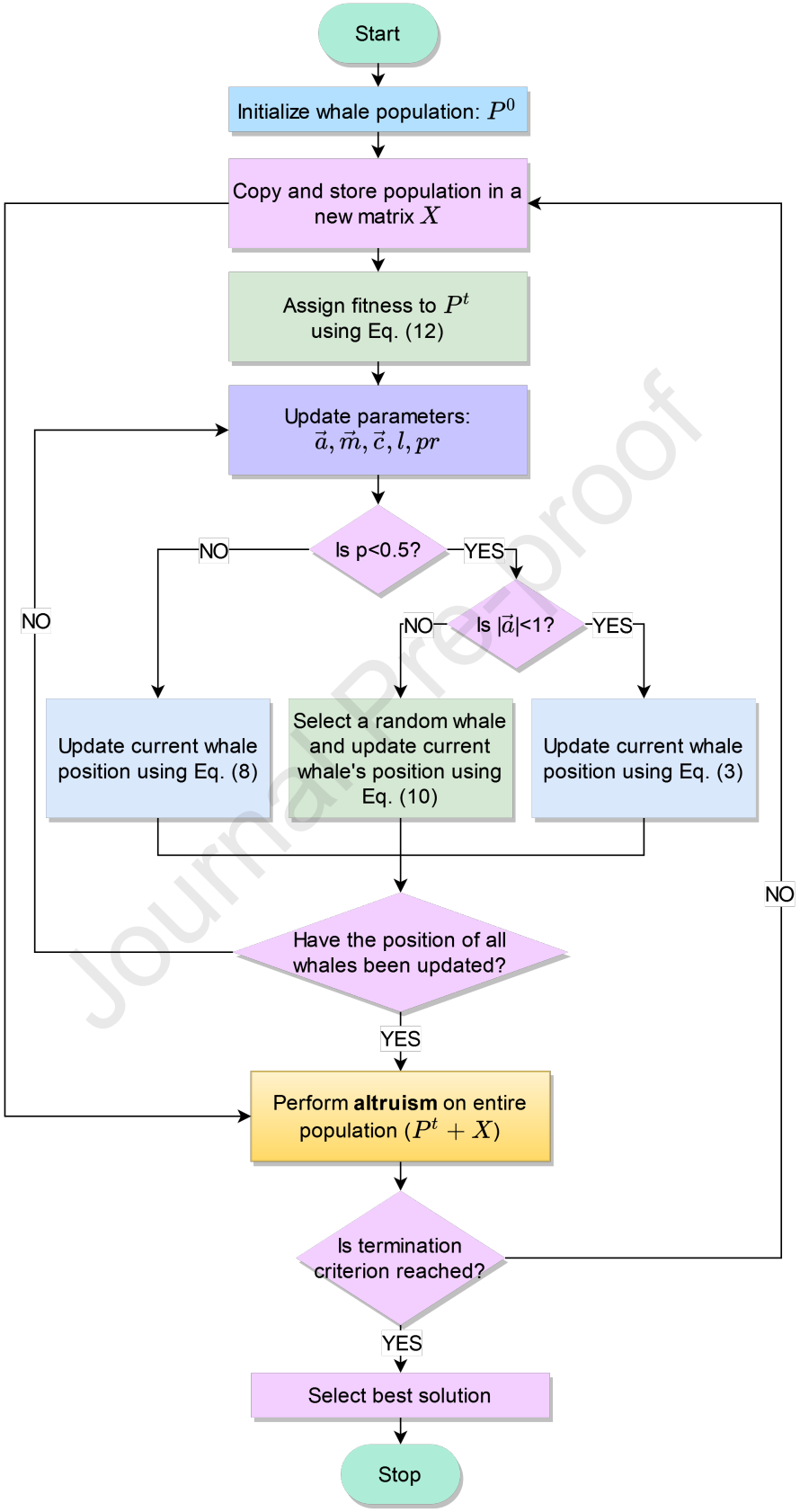


Figure 1: Workflow of the proposed method - AltWOA

above, the filer methods are generally the simplest ones but often fail to outperform wrapper methods. On the other hand, in embedded methods, it is sometimes difficult to find a suitable combination of wrapper and filter methods. Hence in this paper, we propose an improved version of a popular wrapper method called WOA proposed by Mirjalili et al. [4], wherein we embed the concept of altruism in the whale population. The intuition behind using an altruism technique is that for some mediocre fitness solutions, maybe a candidate that is not "fit" enough to be selected in the population pool of the next generation might evolve into a promising solution if allowed to propagate further through the iterations. Thus, a fitter solution might sacrifice itself to allow a candidate with the potential to be an optimal solution in the future to propagate. Furthermore, this study has designed a metric to determine which solution has “potential" and which can be sacrificed, based on the correlation of the features, i.e., a solution with more diverse features selected is allowed to propagate. The proposed method, called Altruistic Whale Optimization Algorithm or AltWOA, shows better predictive capability than most existing state-of-the-art methods when evaluated on eight publicly available gene expression datasets. The highlighting points of the presented research are as follows:

1. With the availability of more and more data in this information technology era, feature selection becomes an important stream of data engineering for optimizing information storage. Our paper addresses this problem by using an evolutionary optimization algorithm by selecting only the relevant features (genes) for classification.
2. In DNA microarray gene expression datasets, the information from thousands of genes is available, but only a handful of them are relevant to identifying the disease. Thus, Feature Selection is critical in this paradigm, allowing clinicians to locate the important genes, the mutations of which cause disease. Such information help researchers devise strategies to prevent and cure diseases occurring from unwanted gene mutations.
3. Meta-heuristics are popularly used for feature selection due to their success in solving several optimization problems. However, according to the “No Free Lunch" theorem, no one algorithm can find the optima in all problems. Thus, we have developed the AltWOA algorithm for addressing the gene subset selection problem in DNA microarray data.
4. The concept of altruism in the traditional WOA algorithm allows candidate solutions, who currently possess mediocre fitness but have the potential to travel to optimal solutions, to thrive through the iterations at the cost of other, less potent solutions.
5. The proposed algorithm is tested on eight high-dimensional microarray datasets and is shown to perform better than existing popular meta-heuristics on the gene selection problem in DNA microarray data, justifying the efficacy of the proposed approach.

The rest of the paper has been organized as follows: Section 2 surveys the literature outlining the development of feature selection methods in gene expression datasets. Section 3 explains in detail the proposed method for feature selection. Section 4 evaluates the proposed AltWOA method on eight popular publicly available gene expression datasets. Finally, Section 5 concludes the findings from this research.

2. Related Work

The ability to acquire a large amount of data in this era is a double-edged sword. While on the one hand, it allows us to analyze features more robustly, on the other hand, storage of such large amounts of data and processing them becomes increasingly difficult. Thus, dimensionality reduction techniques become more critical as it discards irrelevant features without compromising the performance of the learner. Feature subset selection using exhaustive search is an NP-hard problem. Thus several intelligent optimization frameworks have been proposed in the literature to select an optimal subset with far less computational cost. Solving optimization problems using meta-heuristics [8] has become increasingly popular due to its ability to efficiently search for a global optimum. A comprehensive literature survey on the popular metaheuristics used in the literature in various domains of application is tabulated in Table 1.

Every species uses a different strategy in hunting and/or migration. Animals have evolved for millions of years to their current form, where their foraging habits are optimal, leading to a stable population. Thus, studying their behaviors in nature has led to the emergence of optimization algorithms that mathematically try to mimic their schemes. The success of such formulations in a wide range of problems, using considerably fewer resources, has led to their popularity as a research field, where newer strategies are being devised in recent times. Classical optimization algorithms like the Genetic Algorithm (GA) [27] proposed and Particle Swarm Optimization (PSO) [47] have been extensively used in several domains.

However, regardless of the viability of such classical algorithms, they cannot provide optimal solutions for all kinds of optimization problems that researchers need to deal with, as pointed out by the No Free Lunch Theorem [62]. Thus, a range of novel optimization algorithms has also been proposed in the recent literature, mimicking animal behavior, swarm intelligence, or physical phenomena, each having its own set of diverse characteristics. Some recent meta-heuristics based on animal behavior include the Chimp Optimization Algorithm by Khishe et al. [63], and Harris Hawks Optimization by Heidari et al. [64]. Mayfly Algorithm by Zervoudakis et al. [42], and Salp Swarm Algorithm by Mirjalili et al. [65] are swarm intelligence-based algorithms. Archimedes Optimization Algorithm by Hashim et al. [66] and Solar System Optimization by Zitouni et al. [67] are based on physical phenomena. Arithmetic Optimization Algorithm by Abualigah et al. [68], and Sine-Cosine Algorithm by Mirjalili et al. [57] are Mathematics-inspired algorithms. These algorithms have shown their efficacy by achieving benchmark results in various domains of research where optimization is found very crucial. Applications of these algorithms on various such domains are shown in Table 1. Applications of more optimization algorithms of all time

Table 1: Tabular representation of the application domains of different optimization algorithms popularly used in the literature.

Optimization Algorithm	Applications
WOA [4]	1. Kaveh et al. [9]- Sizing optimization problems of truss and frame structures (Mechanical Engineering) 2. Oliva et al. [10]- Parameter estimation in photovoltaic cells (Electrical Engineering) 3. Aljarah et al. [11]- Optimizing weights and biases in ANNs (Computer Science) 4. Abdel et al. [12]- Permutation flow shop scheduling problem (Fluid Mechanics)
BBA [5]	1. Gupta et al. [13]- Classification of white blood cells (Biomedical) 2. Kang et al. [14]- Fault diagnosis for low-speed rolling element bearing failures (Electrical Engineering) 3. Tripathi et al. [15]- Credit scoring of applicants (Finance Industry) 4. Basetti et al. [16]- Optimal Phasor measurement units placement (Power System Monitoring)
CS [17]	1. Yang et al. [18]- Welded beam design problem (Mechanical Engineering) 2. Vazquez et al. [19]- Training Spiking Neural Networks (Artificial Intelligence) 3. Yildiz et al. [20]- Selection of optimal machine parameters in milling operations (Machinery Industry) 4. Tein et al. [21]- Nurse management system (Medical and Hospital Industry)
EO [22]	1. Abdel-Basset et al. [23]- Solar photovoltaic parameter estimation (Electrical Engineering) 2. Wunnava et al. [24]- Multilevel thresholding (Softcomputing) 3. Abdul et al. [25]- Multi dimensions operation of hybrid AC/DC grids (Power Systems) 4. Dey et al. [26]- Speech Emotion Recognition (Signal Processing)
GA [27]	1. Shin et al. [28]- Bankruptcy prediction modeling (Finance Industry) 2. Srivastava et al. [29]- Software Testing (Software Industry) 3. Wang et al. [30]- Calibrating conceptual rainfall-runoff models (Water Resources Research) 4. Norouzi et al. [31]- Optimization of wireless sensor networks (Electronics Industry)
GSA [32]	1. Shaw et al. [33]- Combined economic and emission dispatch problem (Power Systems) 2. Mondal et al. [34]- Minimize the emission of nitrogen oxides and fuel cost (Energy Systems) 3. Hatamlou [35]- Data clustering (Data Mining) 4. Pelusi et al. [36]- Parameter optimization of ANNs (Artificial Intelligence)
HS [37]	1. Lee et al. [38]- Structural design optimization (Mechanical Engineering) 2. Ayvaz et al. [39]- Identification of groundwater parameter structure (Water Management System) 3. Panchal et al. [40]- Therapeutic medical physics (Biomedical) 4. Yazdi et al. [41]- Biped locomotion (Robotics)
MA [42]	1. Kadry et al. [43]- Image multi-level-thresholding (Artificial Intelligence) 2. Wei et al. [44]- Optimization of Fuel Cell (Fluid Mechanics) 3. Hassan et al. [45]- Parameter Estimation of Single-Phase Transformer (Electrical Engineering) 4. Liu et al. [46]- Wind speed forecasting (Energy Systems)
PSO [47]	1. Fan et al. [48]- Load forecasting (Power Systems) 2. Godio et al. [49]- Optimization of electromagnetic geophysical data (Applied Geophysics) 3. Rudek et al. [50]- Skull prosthesis modelling (Medicine) 4. Li et al. [51]- UWB antenna design (Electronics Industry)
RDA [52]	1. Zitar et al. [53]- Optimizing Complex functions (Mathematics) 2. Balashunmugaraja et al. [54]- Provacy preservation in cloud (Business Applications) 3. Nguyen et al. [55]- Pervasive wireless sensor networks (Data Transmission Security) 4. Renuka et al. [56]- Controlling Traffic in Internet of Vehicles (Internet of Things)
SCA [57]	1. Gupta et al. [58]- Train multilayer perceptrons (Artificial Intelligence) 2. Das et al. [59]- Short-term hydrothermal scheduling (Energy Systems) 3. Reddy et al. [60]- Solve profit-based unit commitment problem (Economics) 4. Sahu et al. [61]- Optimized Fractional-order PID controller in photovoltaics (Electrical Engineering)

Journal Pre-proof

Journal Pre-proof

in various domains can be found in the paper of Li et al.[69]. Moreover, some good surveys on these algorithms are also performed such as Yazdani et al.[70] have surveyed evolutionary optimization algorithms, Singh et al. [71] did the same on social media networks, work of Khandija et al. [72] is based on recent advancements of meta-heuristic algorithms and so on.

Hybrid feature selection methods are also being used in the literature to combine the salient features of two or more meta-heuristics. Such algorithms have higher chances of avoiding getting stuck at local optima and exploiting the search space more effectively. Chattopadhyay et al. [73] used an improved Golden Ratio Optimizer, and Yan et al. [74] used a Memetic binary Coral Reef Optimization algorithm embedded with Simulated Annealing (SA) for feature selection. Lopez-Garcia et al. [75] developed the Genetic Algorithm Cross-Entropy (GACE) algorithm, which is an ensemble classification technique [76, 77, 78] based on feature partitioning for imbalanced data classification. Dey et al. [26] hybridized the Golden Ratio Optimizer with the Equilibrium Optimizer, and Al-Thanoon et al. [79] proposed a hybrid of Particle Swarm Optimization with the Dragonfly algorithm. Shukla et al. [80] integrated the Teaching Learning-Based Optimization algorithm with the Gravitational Search Algorithm for feature selection on biomedical datasets. However, hybrid meta-heuristics are computationally much costlier than a single global optimization algorithm, which may be challenging to apply in the practical field with resource constraints.

WOA proposed by Mirjalili et al. [4] is based upon the foraging pattern of humpback whales. Since introduction, it has been successfully applied in several domains like mechanical engineering, electronics, etc. For example, Kaveh et al. [9] used the WOA for optimization of frame and truss problems, Oliva et al. [10] developed an improved WOA for parameter estimation in photovoltaic cells, and Aljarah et al. [11] used the WOA for optimizing connection weights in artificial neural networks. Variants of WOA have also been proposed in the literature to enhance the optimization capability of the basic WOA. For example, Mafarja et al. [81] embedded the Simulated Annealing (SA) as a local search method in the WOA for feature selection applications, Kaur et al. [82] developed the Chaotic WOA, wherein chaotic maps have been embedded in the WOA population to improve its exploratory capability. Bozorgi et al. [83] proposed the Improved WOA, wherein they hybridized the WOA with Differential Evolution for better exploration of the search space. Recently, Gharehchopogh et al. [84] conducted a comprehensive survey on the WOA, its improvements, and its applications. Thus, the scalability of WOA as found in the literature has led us to modify the algorithm further to increase its capabilities and extend it to the gene selection problem in DNA microarray datasets.

Optimization algorithms have been used extensively for feature selection in gene microarray datasets. Baliarsingh et al. [85] developed a Weighted Chaotic Grey Wolf Optimizer method where they used a weighted chaotic map of multiple filter methods to guide the GWO algorithm for gene selection and classification of DNA microarray data. Bommert et al. [86] evaluated the efficacy of 22 filter methods in feature selection and classification of high dimensional microarray data. Baliarsingh et al. [87] proposed a hybrid of SA and Rao algorithm, and Alzaqebah et al. [88] proposed a memory-based Cuckoo Search algorithm for feature selection in gene expression datasets. GÜÇKIRAN et al. [89] evaluated two filter-based feature selection methods- Relief and LASSO coupled with Support Vector Machines (SVM), Multilayer Perceptron, and Random Forest classifiers. Koushik et al. [90] proposed a framework of feature selection on the gene expression datasets using Elastic Net and SVM. In addition to this, Begum et al. [91] have come up with a Fuzzy Preference-Based Rough Set algorithm for biomarkers on gene expression data. Moreover, Ghosh et al. [92] implemented a Recursive Memetic Algorithm for the gene selection task. Almugren et al. [93] conducted a comprehensive survey on hybrid feature selection methods for classification tasks on gene expression datasets. Another experimental study on benchmark microarray data with Filter Ranking Methods is done by Ghosh et al. [94]. Also ensemble feature selection approaches [95, 96, 97, 98] have also become very popular among researchers for detecting various diseases from microarray data. Such as, Abeel et al. [99] developed an ensemble-based feature selection method for cancer detection by identifying the biomarkers wherein they used a recursive procedure and a backward elimination strategy to remove features iteratively. Besides, a unique statistical learning method called the “Resampling method" for the same purpose is proposed by James et al. [100]. Though there are several methods available which experimented on microarray dataset, it is to be noted that proper benchmarking of any algorithms in such datasets is considered as a very important fact. It is because, the efficacy of a software-based approach in Bioinformatics significantly depends on the problem which is being addressed and the dataset chosen for manifesting that. Proving the biological significance of the outcomes of the computer based algorithms is very important to be applied the same in practical purposes. In doing so, various issues arise which are mentioned by Aniba et al. [101]. Also, Mangul et al. [102] showed systematic approaches for benchmarking omics computational tools, whereas Weber et al. [103] mentioned some important guidelines for the same.

Inspired by the need for feature selection in gene expression datasets for efficient storage and computation and by the efficacy of WOA in various optimization problems in different application domains, in this paper, we propose a novel feature selection method, called AltWOA, for the selection of non-redundant features from high dimensional gene expression datasets. The flowchart showing the overall pipeline of the proposed AltWOA algorithm is shown in Figure 1.

3. Proposed Method

This section describes the methodology developed in this study for feature selection in high-dimensional gene expression datasets. First, we describe Pasi Luukka’s filter-based feature ranking algorithm [2] which has been used to select the top 300 genes from the dataset. Then we describe the AltWOA algo-

rithm, which is applied to this reduced feature set to select the optimal feature subset.

3.1. Method of Pasi Luukka

The algorithm of Pasi Luukka et al. [2] is a fuzzy entropy-based feature selection algorithm that is used in this paper for initial dimension reduction of the feature space. After that, the reduced feature set is fed to the AltWOA for further optimization. Pasi Luukka et al. [2] used fuzzy entropy measures and similarity classifiers for feature selection and classification tasks.

The entropy measures can frequently define the variations between the fuzzy set and nicely described crisp set. Now the measures of fuzzy entropy are defined as in Equation 1, where the term $\mu_a(x_i)$ is the i^{th} fuzzy value in the expression.

$$E(a) = -\sum_{i=1}^{n}(\mu_a(x_i)log\mu_a(x_i) + (1-\mu_a(x_i))log(1-\mu_a(x_i))) \quad (1)$$

This fuzzy entropy measure is similar to the measure of fuzziness of the fuzzy set and calculates global deviations from ordinary sets. That is, if a set M_0 is a crisp set, then the entropy of that set would be 0, and the maximum element of ordering can be 0.5.

In this process, one ideal vector vec_i is created, by which the i^{th} class is represented. Although these ideal vectors are generated from sample sets belonging to i^{th} class, these can be user-defined as well. After defining the vectors vec_i, the similarities between test samples and the ideal vectors are calculated. The similarity value determines which class the test sample belongs to. Ideally, suppose a specific test sample is classified to a particular class. In that case, the similarity value between that test sample and the ideal vectors belonging to that class should be equals to 1. Otherwise, the similarity value should be 0. For calculating similarities with the ideal vectors, we get m similarities, where m is the number of features. Here, the fuzzy entropy measures are used for the relevance measurement of the features. Following the entropy calculation (E_1) given in Equation 1, higher entropy values are achieved with similarities close to 0.5, and lower entropy values are obtained while similarities are high. Based on this, the features with the highest entropy are found by calculating the similarities between the ideal feature vector and the sample vector. Thus, the previously mentioned assumption further concludes that the features with higher entropy do not contribute much to deviation among different classes, and features with the lowest entropy values are considered more informative.

For example, let us consider there are 'x' number of samples, 'f' number of features, and 'c' number of classes. After calculating the similarities, the samples are sorted based on the similarity values in the form of a matrix, which has the dimension of $xc \times f$. The fuzzy entropy measures summed by xc values are calculated for each feature in the matrix. After that, the features with the largest fuzzy entropy value are detected and removed from the entire feature set.

3.2. Whale Optimization Algorithm

WOA [4] is a swarm-intelligence-based meta-heuristic algorithm that is developed for addressing complex, continuous optimization problems. As the name suggests, the hunting behavior of humpback whales is the root idea behind this optimization algorithm. Like other nature-inspired optimization algorithms, in WOA, every candidate solution is considered an agent, i.e., whale, which updates its role to reach closer to the prey throughout the iterations. The candidate whales follow certain typical processes to conduct search operations to locate the prey and then attack the prey. In the first method, the whales encircle the prey, and in the second mechanism, the whales create bubble nets to capture the prey. The search space exploration takes place while the whales look for the prey and exploit the hunt area during the attack.

3.2.1. Encircling Prey

In real cases, humpback whales have evocative knowledge about the exact location of the prey. However, inside the multi-dimensional solution space, the location of the ideal candidate solution is not known. Consequently, the final position of the prey (subsequently determined with the aid of the candidate Whales) is taken into consideration as a solution. The iterative process of approaching the global optima goes by elaborating and defining the first and so far the best search agent and, after that, updating the positions of other search agents towards the optima. The following mathematical expressions illustrate the above phenomena.

$$\vec{P}(n+1) = \vec{P}^*(n) - \vec{a} \cdot \vec{d} \quad (2)$$

where $\vec{d}$ is the weighted distance between the position vector of the best solution obtained so far and the current position vector. The mathematical expression of $\vec{d}$ is given as follows-

$$\vec{d} = \mid \vec{c} \cdot \vec{P}^*(n) - \vec{P}(n) \mid \quad (3)$$

In equations Equation 2 and Equation 3, $\vec{P}^*$ is the position vector of the best solution obtained prior to this iteration. Here, the position vector with $(*)$ in superscript denotes the positions of best solutions. The $\vec{a}$ and $\vec{c}$ are the coefficient vectors. It is to be noted that the position, $\vec{P}^*$ gets updated in every iteration in case some better candidate solution is found. The following equations are used to calculate the coefficient vectors $\vec{a}$ and $\vec{c}$.

$$\vec{a} = 2\vec{m} \cdot \vec{R} - \vec{m} \quad (4)$$

$$\vec{c} = 2\vec{R} \quad (5)$$

In Equation 4, the $\vec{m}$, a vector with variable absolute values, decreases from 2 to 0 over the iterations, and the $\vec{R}$ is a random vector that has values within 0 and 1. Now in every iteration, the positions of different candidate solutions around the best solution get updated by appropriate adjustments of the coefficient vectors $\vec{a}$ and $\vec{c}$. The advantage of the random vector $\vec{R}$ is that it can reach any position within the permissible critical points in the search space directly. Therefore, using Equation 2, every position vector can update its positions to get closer to the best solution.

Journal Pre-proof

3.2.2. Bubble-net attacking method

In the Bubble-net attacking method or the exploitation phase, two different approaches are taken to formulate the mathematical model of the attacking. These two approaches are the shrinking encircling mechanism and the spiral updating position.

In the case of the shrinking encircling mechanism, the $\vec{m}$ of Equation 4 is lowered and results in narrows the ranges for $\vec{d}$ as well. In other words, $\vec{d}$ is a random value in the interval of [-m, m], where m is lowered from 2 to 0. The value 'a' can also define the updated position of a search agent within the original position and the position of the best agent of the current iteration.

On the other hand, in the latter mechanism, initially, the distance between the whale (P, Q) and the prey $(P^{'}, Q^{'})$ is calculated. Here position vectors having the $(')$ in superscript denote the locations of prey. The spiral-based equation fitted into the general trajectory of the movement is meant to follow the helix-shaped movement of Humpback whales. The mathematical expression is given below:

$$P(\vec{n+1}) = \vec{D} \cdot e^{bl} \cdot cos(2\pi l) + \vec{P^*}(n) \tag{6}$$

where b and l are constants for the logarithmic spiral shape and a random number lie in the range of [-1, 1] respectively. The $\vec{D}$ is the absolute distance between the prey and the whale, that is

$$\vec{D} = |\vec{P^*}(n) - \vec{P}(n)| \tag{7}$$

The humpback whales simultaneously swim across the prey along a spiral-shaped course, and the movement is also found within a shortening circle. To simulate this behavior, a probability (pr) threshold of 50% is considered for choosing between a spiral model or a shortening encircling mechanism to update the whale's position in the optimization process. Mathematically, it can be represented by Equation 8.

$$P(\vec{n+1}) = \begin{cases} \vec{P^*}(n) - \vec{a} \cdot \vec{d} & if \quad pr < 0.5 \\ \vec{D} \cdot e^{bl} \cdot cos(2\pi l) + \vec{P^*}(n) & if \quad pr \geq 0.5 \end{cases} \tag{8}$$

3.2.3. Exploration with WOA

The variation of $\vec{a}$ vector controls the exploration phase of the WOA. Due to the uncertain position of the whales, the search is a bit random. Therefore, $\vec{a}$ can be used with random values greater than 1 or less than 1 for mobilizing search agent on the way to circulate some distance away from a reference whale. Unlike the exploitation phase, in the exploration phase, the position of the search agent is upgraded based on a randomly selected search agent and not the best agent. Mathematically,

$$\vec{D} = |\vec{P_r} - \vec{P}| \tag{9}$$

and,

$$\vec{P}(n+1) = \vec{P_r} - \vec{a} \cdot \vec{d} \tag{10}$$

where P_r is the randomly selected search agent, chosen for upgradation of the position of the current search agent.

3.2.4. WOA for Feature Selection

It is to be noted that WOAis proposed to solve continuous optimization problems. Therefore, to convert the algorithm and make it suitable for feature selection tasks, it is needed to map the continuous search of WOA to binary search. A Sigmoid function is used as a transfer function which is given by-

$$F(m) = \frac{1}{1 + e^{-m}} \tag{11}$$

Additionally, the fitness function used to evaluate a candidate solution 'C' of dimension 'D' in the population is given by Equation 12, where '$acc(C)$' is the classification accuracy for the candidate solution and '$FS(C)$' is the number of features selected by the candidate solution C. The fitness function is the weighted sum of the accuracy and the fraction of features *not* selected by the candidate solution. Thus the optimization problem is the maximization of the fitness function. The value of the weight α is set as 0.7 through extensive experiments in this study.

$$fitness(C) = \alpha \times acc(C) + (1 - \alpha) \times \frac{D - FS(C)}{D} \tag{12}$$

3.3. Altruism

Altruism means sacrificing oneself for the sake of others. In the context of an optimization algorithm, some of the individuals in the population will be sacrificed to give a chance for evolution to some other individuals, which, otherwise, would have been eliminated due to lower fitness. The intuition behind using such a technique is that there may be cases where an individual who currently has a decent fitness value has the potential to evolve into a promising solution through the iterations if given a chance without eliminating it through elitism. However, to give such a solution a chance to evolve, another individual with a fitness value close or same as the prior needs to be dropped from the solution pool through some evaluation criteria, for example, checking the average correlation of the elements selected by the two individuals. The proposed altruism scheme used such an approach which is explained below.

For computing the final population after the altruism operation, the whole pool of individuals is taken as input, i.e., the individuals from the previous generation and new individuals generated through the WOA evolutionary operations are together taken as input. So, if the population size of the WOA is set as N, then the input to the altruism operator becomes $2N$. In other words, the elitism operation, where the N most fit individuals from this $2N$ sized population pool were selected, is now replaced by the altruism operator. The number of altruistic individuals, i.e., the number of individuals to be sacrificed for the benefit of others, is pre-set as a hyperparameter to the proposed AltWOA method. Let A denote that value.

Since the altruism operation is used on individuals with medium-ranged fitness values, a set of best-fit solutions is kept intact (i.e., no altruism is applied). The number of such solutions is $I = N - A$.

Now, from the remaining $2N - I$ individuals, the top $2A$ number of individuals is considered. The primary altruism function

will be applied to these $2A$ individuals. We will evaluate these solutions in pairs, and after the altruism operation, we will select only A number of individuals from these $2A$ individuals. So, previously, I solutions were kept intact, and now A solutions are selected. Therefore, the final population size becomes $I + A = N$, and thus the algorithm is coherent.

Let us now consider the operation which will select A individuals from the mediocre set of $2A$ solutions. The $2A$ solutions are first grouped in pairs: each solution is selected and paired up with the most "similar" solution. The similarity index (SI) between a pair of candidate solutions C_1 and C_2 having fitness values A_1 and A_2 respectively is defined by Equation 13, where HD denotes the Hamming distance between C_1 and C_2. The Hamming distance $HD(C_1, C_2)$ between the two binary encoded candidate solutions is the number of positions where the elements are different, i.e., positions where C_1 is 1, C_2 is 0 and vice versa. ϵ_1 and ϵ_2 are minimal values added to the denominator to handle cases where the denominator becomes 0 giving an error. The proposed work sets the weighting factor β experimentally to 0.3.

$$SI(\beta, C_1, C_2, A_1, A_2) = \frac{\beta}{HD(C_1, C_2) + \epsilon_1} + \frac{1-\beta}{|A_1 - A_2| + \epsilon_2} \quad (13)$$

The two candidate solutions are more closely related when the SI value is high. Thus, concerning the solution C_1, the paired solution C_2 is selected as the one with the highest similarity index from the pool of $(2A - I)$ solutions. Now, out of these two solutions, one will be sacrificed for the sake of the other. The criteria for selecting the sacrificial solution is defined by a correlation score metric (CSM_m) for the candidate solution C_m as shown in Equation 14.

$$CSM_m = \alpha \times SCC_{avg}(C_m) + (1-\alpha) \times PCC_{avg}(C_m) \quad (14)$$

SCC_{avg} and PCC_{avg} represent the average value of Spearman's Correlation Coefficient (SCC) and Pearson's Correlation Coefficient (PCC) for the features selected. The SCC and PCC values for the entire feature set are computed at the offset of the algorithm, each of which is a $D \times D$ matrix, where D represents the dimensionality of the feature set. Now, for computing the $SCC_{avg}(C_m)$ for candidate solution C_m which has, say, d number of features selected, the corresponding $d \times d$ matrix are sifted out from the entire SCC matrix, i.e., the SCC values for the features that have been selected in the C_m solution forms this $d \times d$ matrix. Then the average value of this matrix is calculated, which is denoted by $SCC_{avg}(C_m)$. Following a similar workflow, the $PCC_{avg}(C_m)$ is computed. Thus, these two values are summed with weights to generate the correlation score CSM_m. A lower correlation score identifies the candidate solution to have the potential for improvement through evolution. This is because a smaller correlation value is indicative of higher feature importance, and a high correlation represents redundancy. Thus, between a pair of candidate solutions, say C_1 and C_2, the one with a lower CSM value is selected to be in the population pool, and the other is eliminated.

PCC for feature x and label y, denoted by PCC_x is given by Equation 15, where $cov(X, Y)$ represents the covariance between feature set X and labels Y, and the variances of distributions X and Y are given by σ_x and σ_y respectively.

$$PCC_x = \frac{cov(X, Y)}{\sigma_x \sigma_y} \quad (15)$$

SCC is a non-parametric measure of the rank correlation between variables and is defined as the PCC between rank variables. SCC tries to assess how well a monotonic function can be used to describe the relationship between two variables. Unlike the Pearson's correlation which assesses the linear relationship, the SCC analyses the monotonic relationship between variables. For a feature set of size n, the attributes X_i, and labels Y_i are converted to ranks r_{X_i}, r_{Y_i} and the Spearman rank R_S is computed by Equation 16, where $cov(r_{X_i}, r_{Y_i})$ represents the covariance of r_{X_i} and r_{Y_i}, and the standard deviations of r_{X_i} and r_{Y_i} are denoted respectively by σ_{r_X} and σ_{r_Y}.

$$R_S = \frac{cov(r_X, r_Y)}{\sigma_{r_X} \sigma_{r_Y}} \quad (16)$$

Despite the apparent similarities between the PCC and SCC measures, there are some fundamental differences. An SCC of 1 is obtained when the two comparable attributes, whose relationship may not be linear, are monotonically related. However, this does not ensure a PCC of 1. SCC is less sensitive to strong outlier data points than PCC since the covariance limits the outlier value to its rank in the SCC computation. SCC and PCC only perform similarly when the data under consideration have no prominent outliers and roughly follow an elliptical distribution in the feature space. Such cases are ideal and do not generally occur in real-world datasets. Thus, since the PCC and SCC compute the correlation between variables in distinct ways, the weighted average of these two measures is a robust computation of the correlation between the features selected and thus has been selected in the proposed method.

The flowchart of the AltWOA is shown in Figure 1. Note that the CSM_m value is not dependent on the classification accuracy or the number of features selected by the candidate solution (and thus independent of the fitness parameter, which is a weighted sum of accuracy and fraction of features selected). It is only determined by the intrinsic properties of the features that have been selected by the candidate solution C_m. Thus an individual with a higher fitness value may get discarded due to a high value of CSM_m. The pseudo-code for the altruism operation is shown in Algorithm 1.

3.4. Computational complexity

In general, it is considered that wrapper-based feature selection algorithms have the highest computational complexity compared to other types of feature selection algorithms. Therefore, it is crucial to be aware of the computational complexity while proposing a wrapper feature selection algorithm.

In the case of AltWOA, there are three parameters which are to be considered for computing the computational complexity of the algorithm. These are: number of iterations (i), population

Journal Pre-proof

Algorithm 1 Pseudo code for the Altruism operation

Input:
Population P of dimension $2N \times D$
SCC score matrix "SCC" of dimension $D \times D$
PCC score matrix "PCC" of dimension $D \times D$
Number of altruistic individuals: A
Output:
Final population P_{alt} of dimension $N \times D$

```
Initialize empty N × D matrix for P_alt
Add best-fit I = N − A individuals in P_alt
Select the next 2A individuals from the altruism pool.
Group pairs of individuals according to similarity index (Equation 13)

for every group pair do
    Calculate CSM_1 and CSM_2 for the two individuals C_1 and C_2 in the current group pair using Equation 14

    if CSM_1 ≤ CSM_2 then
        Add C_1 to P_alt

    else
        Add C_2 to P_alt
    end if
end for
```

size (p) and number of altruistic individuals (a). Now in each iteration, algorithm iterates over the population size and the altruistic individuals for once only, therefore the complexity in one iteration (it) is

$$O(it) = O(p) + O(a) \tag{17}$$

Therefore, for a total of i number of iterations the total computational complexity of the AltWOA algorithm would be

$$O(A) = O(i) \cdot O(it) \tag{18}$$

that is,

$$O(A) = \Omega(i) \cdot [O(p) + O(a)] \tag{19}$$

4. Results and Discussion

This section assesses the proposed framework by performing several experiments and comparisons with some existing methods. To estimate the reliability of the proposed AltWOA method more robustly, we report accuracy values obtained from the method on several datasets and evaluate the model based on other evaluating parameters such as precision, recall, and F1-score. The mathematical expressions of the mentioned evaluation metrics for multi-class classification problem are given by Equations 20, 21, 22 and 23, where M_{ij} is an element of confusion matrix at i^{th} row and j^{th} column.

$$Accuracy_i = \frac{\sum_i M_{ii}}{\sum_i \sum_j M_{ij}} \tag{20}$$

$$Precision_i = \frac{\sum_i M_{ii}}{\sum_i \sum_j M_{ji}} \tag{21}$$

$$Recall_i = \frac{\sum_i M_{ii}}{\sum_j M_{ij}} \tag{22}$$

$$F1 - Score_i = \frac{2}{\frac{1}{Precision_i} + \frac{1}{Recall_i}} \tag{23}$$

To assess the performance of a model thoroughly, all of the accuracy, precision, recall, and F1-score play a significant role. The accuracy gives the overall idea about the performance of a mathematical model, whereas, for class-wise performance, precision and recall are essential. Moreover, the accuracy often fails to estimate a model's performance while significant class imbalance is present in the dataset. Therefore the F1-score is taken into account.

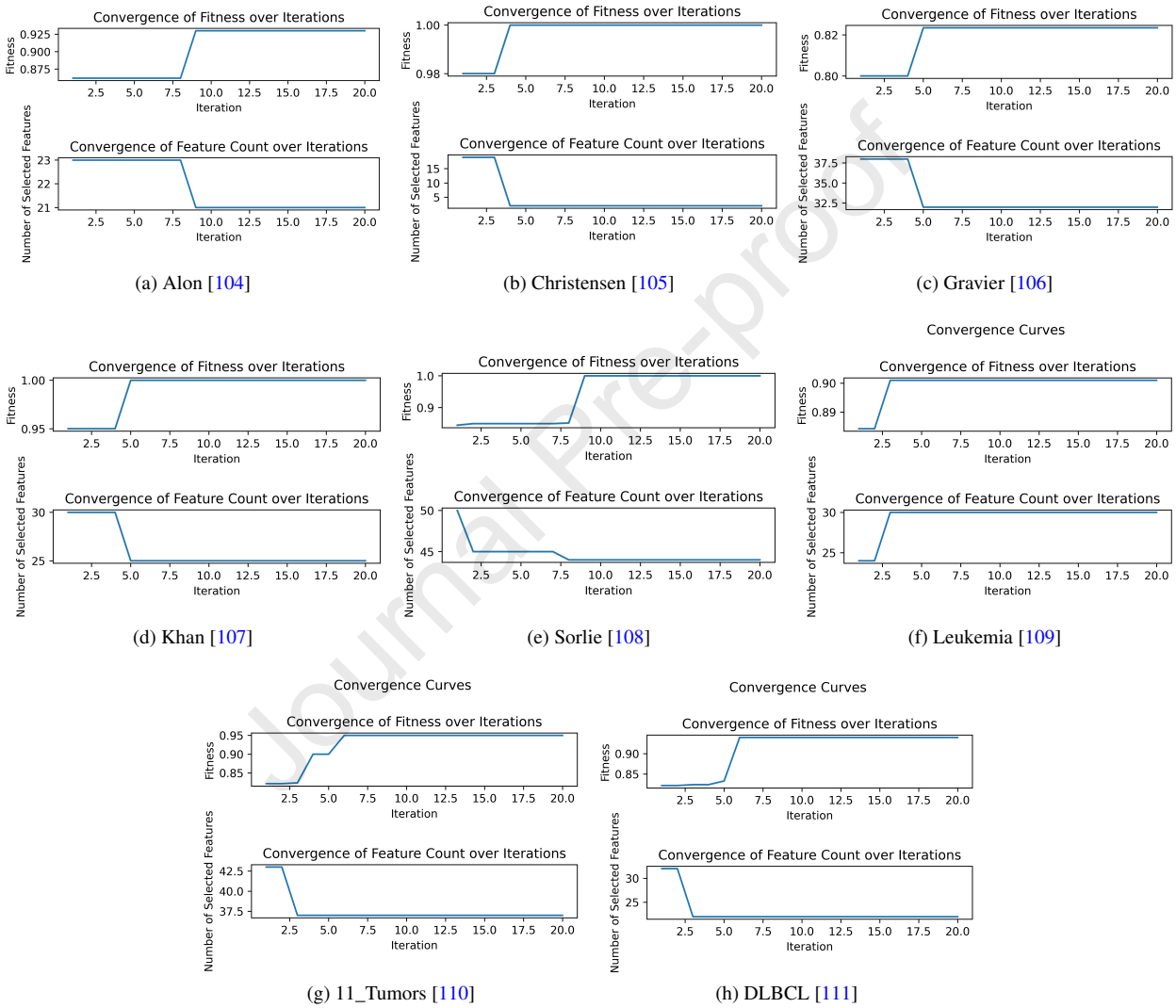
4.1. Description of Datasets Used

To analyze the performance of the proposed method, we use eight high dimensional DNA microarray gene expression datasets, both binary class and multi-class. The description of the datasets is provided in Table 2. Except for the dataset by Christensen et al. [105], Leukemia [109] and DLBCL [111] the other five datasets are used for detecting cancer. The Christensen [105] dataset is a general-purpose gene expression dataset used to capture any kind of abnormalities, the Leukemia [109] dataset consists of genes for identifying leukemia, and the DLBCL (Diffuse Large B-Cell Lymphoma) [111] dataset is for identifying lymphoma. Alon [104], Gravier [106], and DLBCL [111] datasets are binary class, while Christensen [105], Khan [107] Sorlie [108], Leukemia [109], and 11_Tumors[110] are multi-class datasets. Evaluation of these varied datasets gives a more robust measure of the efficacy of the proposed feature selection approach.

4.2. Implementation

This section discusses the results obtained by performing various experiments using AltWOA on the eight different microarray datasets, which eventually comprehensively estimates the classification ability, robustness, and reliability of AltWOA. To achieve that, the meta-heuristic algorithm AltWOA is evaluated on a scale of different evaluation metrics such as accuracy, precision, recall, and F1-score. In this work, to avoid any biasness in the dataset which might occur during the feature selection process of high-dimensional data [112] we have performed 5-fold cross-validation on each of the mentioned datasets, and all the results reported are the average of 5 folds. In this case, we should note that all the experiments of different folds are mutually exclusive and independent of each other for a particular dataset. Therefore, there is no overlap or dependency between two different folds of a dataset.

Moreover, the test data remains completely unseen to the AltWOA during the whole feature selection process. That is, during the iterative evaluation of the AltWOA, it deals with only the training data. At the end of the feature selection process,



(a) Alon [104] (b) Christensen [105] (c) Gravier [106]

(d) Khan [107] (e) Sorlie [108] (f) Leukemia [109]

(g) 11_Tumors [110] (h) DLBCL [111]

Figure 2: Convergence plots obtained by the proposed AltWOA feature selection algorithm on the microarray gene expression datasets used in this study: (a)Alon [104] (b) Christensen [105] (c) Gravier [106] (d) Khan [107] (e) Sorlie [108] (f) Leukemia [109] (g) 11_Tumors [110] and (h) DLBCL [111].

Journal Pre-proof

Table 2: Description of the DNA microarray datasets used in this study. SRBCT: Small Round Blue Cell Tumor, DLBCL: Diffuse Large B-Cell Lymphoma

Dataset	Samples	Genes	Classes	Disease
Alon [104]	62	2000	2	Colon Cancer
Christensen [105]	217	1413	3	N/A
Gravier [106]	168	2905	2	Breast Cancer
Khan [107]	63	2308	4	SRBCT (Cancer)
Sorlie [108]	85	456	5	Breast Cancer
Leukemia [109]	72	7129	2	Leukemia
11_Tumors [110]	174	12534	11	Carcinoma
DLBCL [111]	77	5470	2	Lymphoma

Table 3: Results obtained by the proposed method on the gene expression microarray datasets used in this study.

Dataset	Accuracy	Precision	Recall	F1-Score	No. of Features Selected
Alon [104]	1.0000	1.0000	1.0000	1.0000	21
Christensen [105]	1.0000	1.0000	1.0000	1.0000	2
Gravier [106]	0.9412	0.9510	0.9412	0.9461	44
Khan [107]	1.0000	1.0000	1.0000	1.0000	25
Sorlie [108]	1.0000	1.0000	1.0000	1.0000	35
Leukemia [109]	1.0000	1.0000	1.0000	1.0000	30
11_Tumors [110]	1.0000	1.0000	1.0000	1.0000	37
DLBCL [111]	1.0000	1.0000	1.0000	1.0000	22

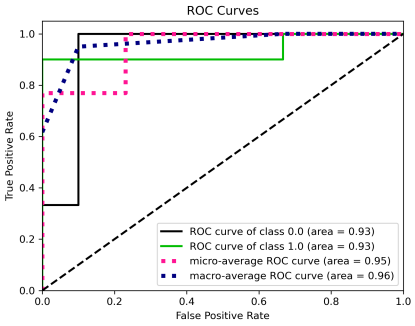
the AltWOA outputs the binary encoded candidate solution possessing the best fitness value. After that, the original feature set is reduced to the lower-dimensional optimal feature subset. The classifier (SVM in our case) is trained on the training samples with the feature subset selected by AltWOA and then evaluated on the test data. The results obtained by the AltWOA algorithm on the eight datasets are depicted in Table 3. We can observe from the table that impressive results are obtained on all eight datasets using the proposed framework. Except for Gravier, 100% is obtained for all four evaluation metrics on other datasets. Besides, the number of genes selected using this framework is also significantly less. The convergence plots obtained by the proposed AltWOA algorithm are shown in Figure 2.

Moreover, for further estimation, we also report the ROC curves obtained by AltWOA on all eight datasets as Figure 3. It is evident that except for the 11_Tumors and the Alon datasets, Area Under the Curves (AUCs) of all classes of all other datasets are either exactly 1.0 or very near to 1.0. For the 11_Tumors dataset, 7 out of 10 classes have high AUC, and only three classes report moderate AUC resulting in quite promising micro-average ROC of 90% and macro-average ROC of 92%. Similar results are observed for the Alon dataset as well. Though the AUCs are not exactly 100% but are very near to that. Therefore, the overall performance of the proposed algorithm based on ROC curves can be considered quite promising. Hence, obtained results confirm the reliable performance of the AltWOA on all of the datasets considered in this study for evaluation.

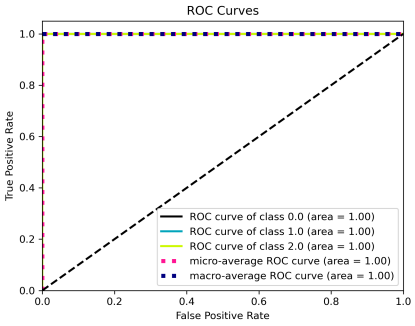
It is to be noted that the proposed method is implemented using the Python3 programming language with several data science and machine learning-based in-built packages available. The most significant packages among them are *Scikit-learn*, *Pandas*, and *Numpy*. Among these, *Pandas* is used to read the data from table form (in our case, it is a *.csv*) file, and *Numpy* is used for regulating the flow of data over generation strictly in array format. The role *Scikit-learn* is versatile– from calculating the fitness of a candidate solution to obtaining final confusion matrices, every statistical and predictive task has been performed with the aid of this package. The codes have been executed on a machine having 8 GB RAM and Intel i5, the processor with Operating System as Windows.

4.3. Hyperparameters Tuning

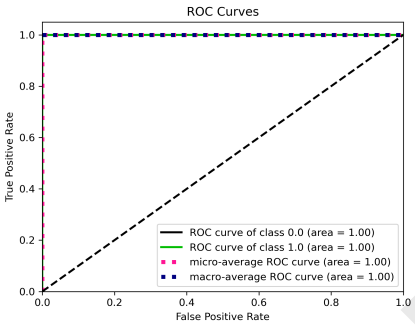
The multi-staged algorithmic complexity of the proposed AltWOA method consists of some essential hyperparameters that affect the classification results noticeably. One such hyperparameter is the number of altruistic individuals of the algorithm, i.e., the number of candidate solutions that go through altruism to generate a more prolific population. While performing several experiments, this hyperparameter comes out to be very important since the slight variation of this parameter causes significant changes in the results. The variation of classification accuracies, fitness, and percentage of features selected concerning the number of altruistic individuals on all eight microarray datasets are depicted in Figure 4. We can observe that the variation in fitness and accuracies show a pattern. Initially, it increases, and at a particular maximum value is achieved, and after that, it starts decreasing and flattens after some point. Just the opposite is observed for the case of the number of features selected. Therefore, this indicates that the optimal value of this hyperparameter is somewhere between left and right extremums, and for all of the cases, this is found to be 10%.



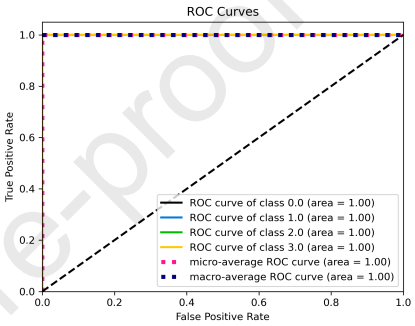
(a) Alon [104]



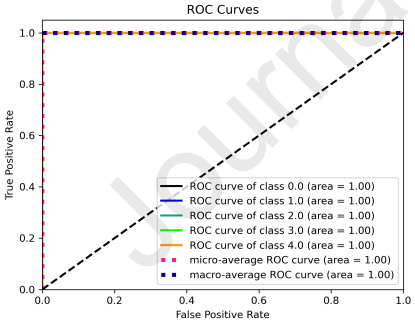
(b) Christensen [105]



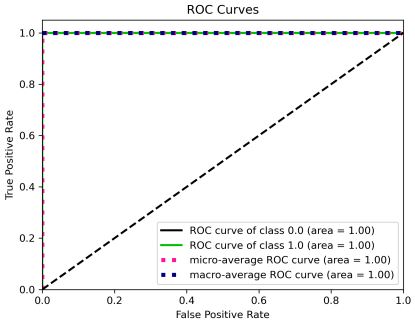
(c) Gravier [106]



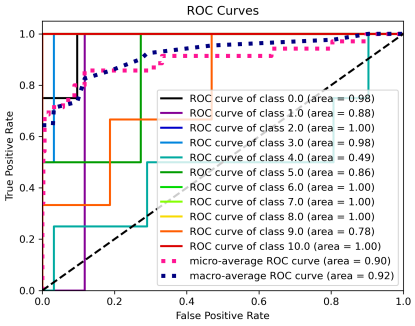
(d) Khan [107]



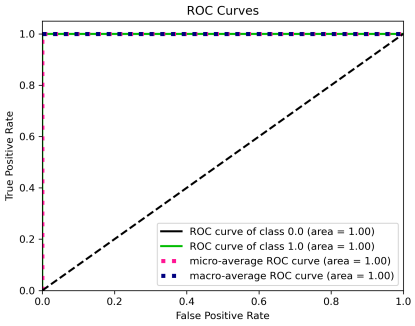
(e) Sorlie [108]



(f) Leukemia [109]



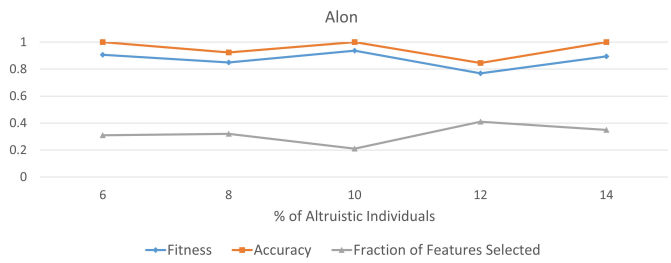
(g) 11_Tumors [110]



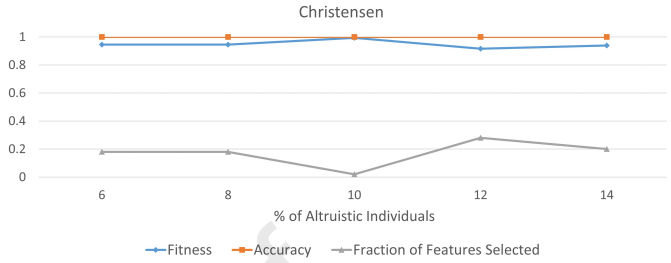
(h) DLBCL [111]

Figure 3: ROC curves obtained by the proposed AltWOA method on the datasets used in this study: (a)Alon [104] (b) Christensen [105] (c) Gravier [106] (d) Khan [107] (e) Sorlie [108] (f) Leukemia [109] (g) 11_Tumors [110] and (h) DLBCL [111].

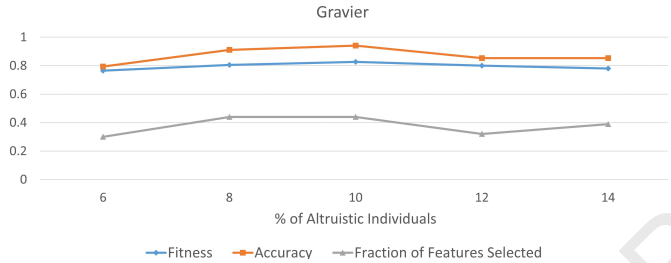
Journal Pre-proof



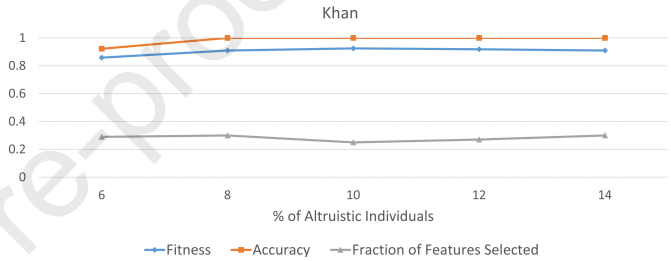
(a) Alon [104]



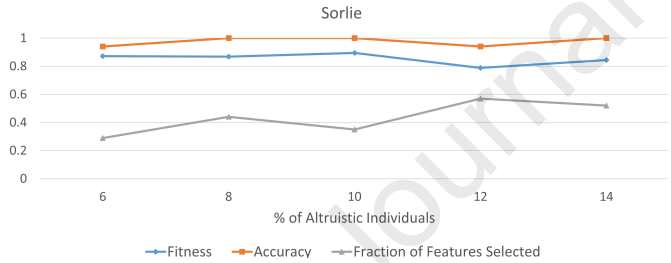
(b) Christensen [105]



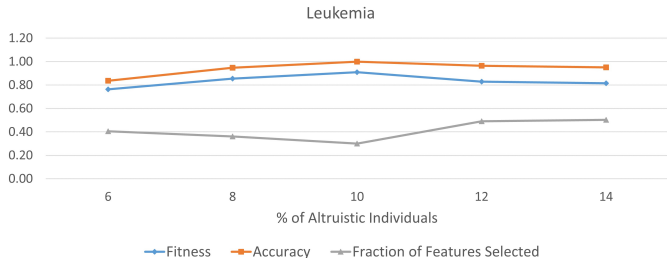
(c) Gravier [106]



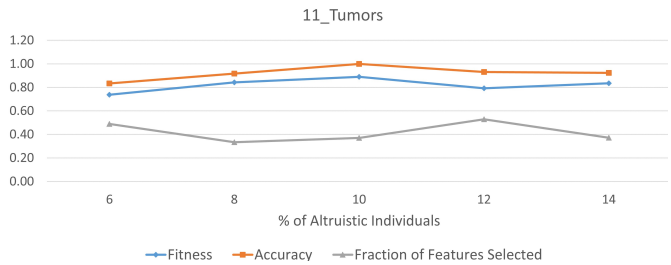
(d) Khan [107]



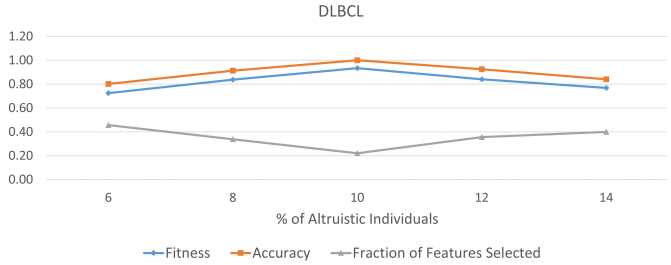
(e) Sorlie [108]



(f) Leukemia [109]



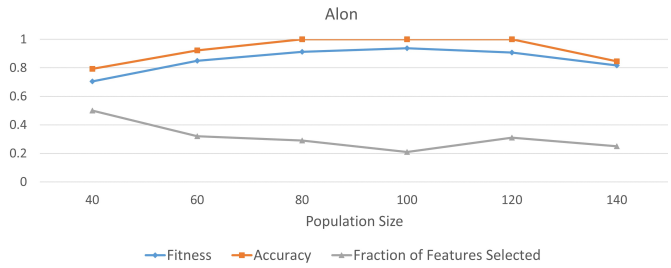
(g) 11_Tumors [110]



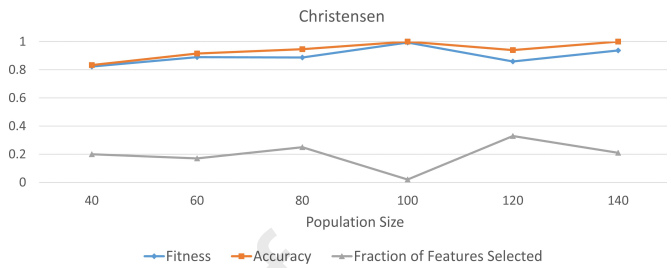
(h) DLBCL [111]

Figure 4: Performance comparison with respect to the percentage of altruistic individuals (a hyperparameter) in the AltWOA on the eight gene expression datasets: (a)Alon [104] (b) Christensen [105] (c) Gravier [106] (d) Khan [107] (e) Sorlie [108] (f) Leukemia [109] (g) 11_Tumors [110] and (h) DLBCL [111]. Population size for the experiments here have been set to 100.

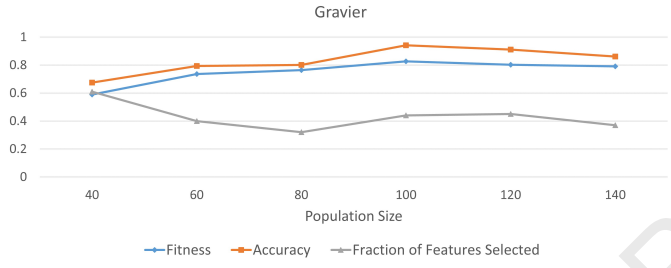
Journal Pre-proof



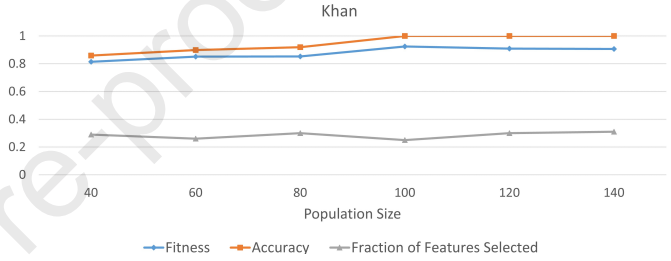
(a) Alon [104]



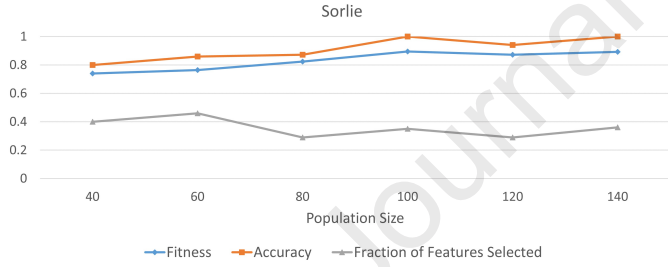
(b) Christensen [105]



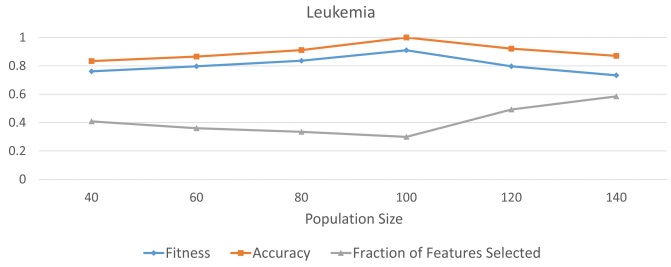
(c) Gravier [106]



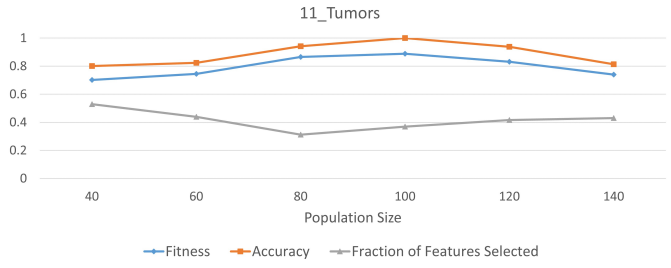
(d) Khan [107]



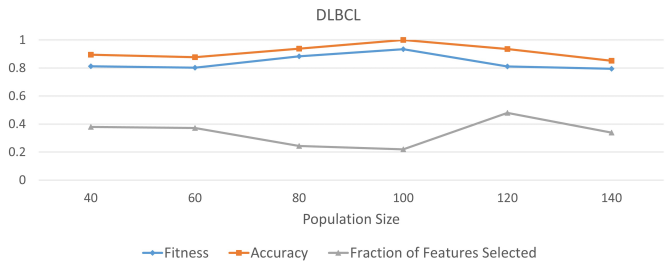
(e) Sorlie [108]



(f) Leukemia [109]



(g) 11_Tumors [110]



(h) DLBCL [111]

Figure 5: Performance comparison with respect to the population size in the AltWOA on the eight gene expression datasets: (a)Alon [104] (b) Christensen [105] (c) Gravier [106] (d) Khan [107] (e) Sorlie [108] (f) Leukemia [109] (g) 11_Tumors [110] and (h) DLBCL [111]. Percentage of altruistic individuals for the experiments here have been set to 10%.

Journal Pre-proof

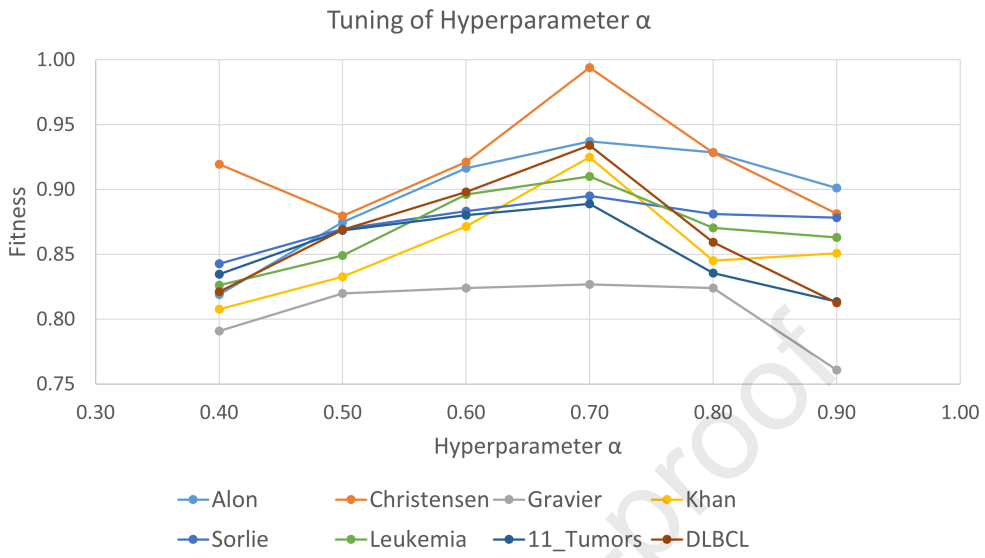


Figure 6: Comparison of performance of the proposed algorithm by varying the hyperparameter α used in Equation 12.

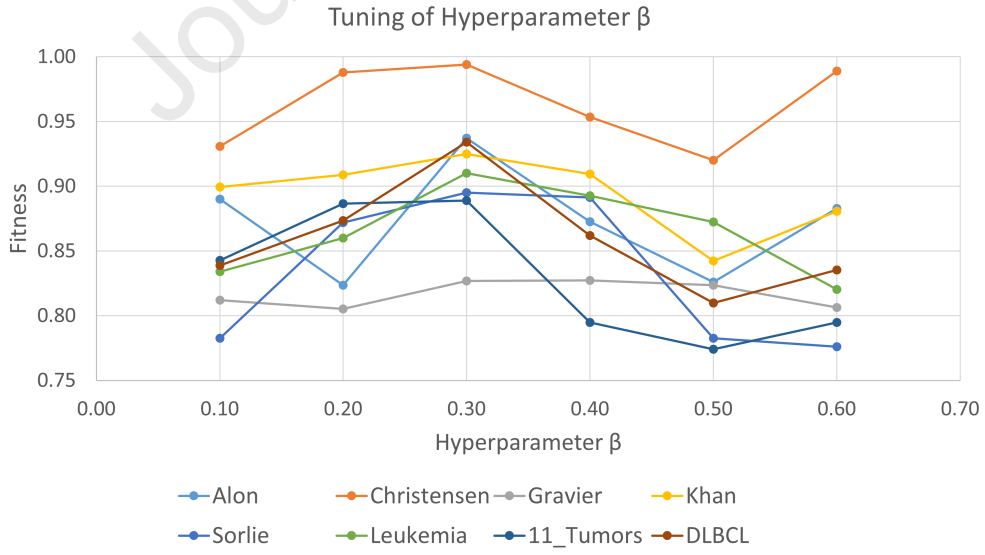


Figure 7: Comparison of performance of the proposed algorithm by varying the hyperparameter β used in Equation 13.

Journal Pre-proof

Another significant hyperparameter is the number of randomized populations initialized before the optimization begins, and this is a critical hyperparameter for all optimization algorithms. The Figure 5 shows the variation of the accuracy, fitness, and the number of features selected for final classification on all eight datasets. We can observe that we can find no crucial pattern from the resultant plots, but the variation can be seen clearly. However, we can see that an initial population size of 100 produces good results in all three parameters for all the datasets. Therefore, to maintain uniformity, we have considered the initial population of the AltWOA as 100, and for all further experiments discussed next, the initial population is fixed at a value of 100.

In addition to these, in this robust study, values of two more hyperparameters are also determined by performing exhaustive experimentation. These two hyperparameters are the α used in the fitness equation in Equation 12 and the β used in the calculation of the similarity index in Equation 13. Figure 6 shows the comparison of the performance by varying the hyperparameter α. We can note that the performance peaks at $\alpha = 0.7$, and thus, this value has been set for all experiments. Similarly, Figure 7 shows the comparison of the performance by varying the hyperparameter β. In this case, we can note that the performance peaks at $\beta = 0.3$, and thus, it has been set as the value of β for all experiments.

4.4. Comparison with state-of-the-art

In order to assess the classification ability of AltWOA in a more robust way, some popular wrapper-based feature selection algorithms have also been evaluated on the mentioned datasets and a comparative study is made based on the results. The algorithms which have been chosen for comparison are:

1. Binary Bat Algorithm (BBA) by Mirjalili et al. [5]
2. Cuckoo Search (CS) by Yang et al. [17]
3. Equilibrium Optimizer (EO) by Faramarzi et al. [22]
4. Genetic Algorithm (GA) by Holland et al. [27]
5. Gravitational Search Algorithm (GSA) by Rashedi et al. [32]
6. Harmony Search (HS) by Geem et al. [37]
7. Mayfly Algorithm (MA) by Zervoudakis et al. [42]
8. Particle Swarm Optimization (PSO) by Kennedy et al. [47]
9. Red Deer Algorithm (RDA) by Fathollahi et al. [52]
10. Sine-Cosine Algorithm (SCA) by Mirjalili et al. [57]
11. Basic WOA by Mirjalili et al. [4]

It is to be noted that the Optimization Algorithms (OAs) choices are not made randomly. The GA, HS, and PSO are old though prevalent optimization algorithms, which can compete with any optimization algorithms of recent times. On the other hand, the rest of the algorithms are comparatively recently developed and have become very popular due to their outstanding performance in various domains. The classification accuracy and number of features selected for the final feature set of all these algorithms along with AltWOA are reported in Figure 8. Note that, for all of these optimization algorithms, the ratio of initial feature compression using Pasi Luukka algorithms has been made fixed such that fair comparative experiments can be conducted among them.

Results reported in Figure 8 clearly show that the proposed AltWOA not only achieves the highest classification accuracies among all the other algorithms but also uses the lowest number of features for final classification in all eight gene expression microarray datasets. Along with AltWOA, the GA, the PSO, and the EO perform better than the other algorithms. On average, although the accuracies achieved by GA are sufficiently high, the number of features selected for final classification is also noticeably higher than that of others. The other algorithms also achieved promising results except for the RDA and the CS. SCA is found to be achieving outstanding classification accuracy. Although similar to GA, the dimension of the final feature set obtained from SCA is significantly larger than AltWOA. Therefore, considering the overall performances of these algorithms, we can conclude that, AltWOA performs superior to other algorithms chosen for comparison.

4.5. Convergence plots

The convergence plots are considered reasonable measures to estimate the performance of an optimization algorithm. In this study, we have reported two types of convergence plots on all eight datasets. One is the fitness variation over the iterations, and the other is the variation of the number of features selected over the iterations.

As mentioned in the previous section, the fitness of a candidate solution depends on the accuracy and the number of features selected by that particular candidate solution. A higher fitness value indicates a better candidate solution. On the other hand, the primary objective of the algorithm is to reduce the feature set dimension for a particular dataset. Therefore, ideally, the fitness should increase over the iterations, and the number of features selected should decrease over the iterative learning process. Although, in reality, the monotonically increasing curve is never obtained, instead the curve gets saturated after a certain number of iterations. The point of saturation can vary from dataset to dataset for a particular algorithm. Similar characteristics are observed from the convergence plots of AltWOA while evaluated on the datasets used in this study. The convergence plots obtained by the AltWOA methods on all eight microarray datasets are shown in Figure 2. As a whole, it can be observed that the AltWOA method follows the ideal characteristic, and for all the cases, the saturation is obtained very fast (within ten iterations). Therefore, it ensures the fast converging capability of the proposed framework, which reduces the computation cost of the overall method.

4.6. Computation Time

Since we have proposed a wrapper-based feature selection algorithm, it is imperative to observe the computation time of the algorithm. We have performed a comparative study of the total computation times taken by the different feature selection algorithms on 20 iterations. The times taken by all the metaheuristics on the eight datasets used in this study are reported in

(a) Alon [104]

(b) Christensen [105]

(c) Gravier [106]

(d) Khan [107]

(e) Sorlie [108]

(f) Leukemia [109]

(g) 11_Tumors [110]

(h) DLBCL [111]

Figure 8: Comparison of different optimization algorithms for feature selection on the microarray datasets used in this study: (a)Alon [104] (b) Christensen [105] (c) Gravier [106] (d) Khan [107] (e) Sorlie [108] (f) Leukemia [109] (g) 11_Tumors [110] and (h) DLBCL [111].

Table 4: Comparison of the computation time (in seconds) of the AltWOA and popular meta-heuristics in literature, over 20 iterations.

Dataset	**Computation Time (s)**											
	AltWOA	**WOA**	**BBA**	**CS**	**EO**	**GA**	**GSA**	**HS**	**MA**	**PSO**	**RDA**	**SCA**
Alon [104]	10.37	9.31	9.71	10.67	12.23	11.53	11.34	9.88	12.20	10.12	9.72	8.34
Christensen [105]	19.36	18.44	18.80	21.12	15.47	15.05	18.35	21.76	18.86	20.64	16.88	17.60
Gravier [106]	23.63	18.03	25.30	25.81	24.53	20.59	26.43	26.29	26.17	20.01	19.05	23.16
Khan [107]	20.45	16.51	11.15	15.94	14.23	14.02	11.03	15.88	15.02	11.92	14.06	12.86
Sorlie [108]	13.42	9.60	11.55	13.25	13.85	15.53	11.16	13.49	13.98	14.16	12.17	11.38
Leukemia [109]	41.53	39.39	38.35	39.21	38.17	38.06	38.06	39.90	39.87	39.54	38.54	38.44
11_Tumors [110]	262.98	260.01	271.47	259.62	250.24	262.35	253.41	267.78	261.23	252.15	255.46	257.91
DLBCL [111]	563.02	561.34	564.35	563.29	564.25	566.59	563.42	568.82	568.50	567.31	565.04	566.83

Table 4. We can observe that the time taken by the AltWOA is slightly higher than some of the fast converging algorithms such as WOA, RDA, and SCA. The extra altruism operation running on top of the traditional WOA takes this excess time. However, the difference between the total computation time over the 20 iterations is not significant, although the AltWOA performs significantly better than the traditional WOA algorithm, making this small margin of increase in computation time worthwhile.

4.7. Feature Selection Stability

Stability in feature selection [113, 114, 96] has become a vital parameter measuring the strength of optimization algorithms in recent times. The stability of the genes selected by optimization algorithms may help clinicians for better diagnosis as such genes possess some unique characteristics due to which the algorithms select these. For example, in the case of cancer detection from microarray gene expression data, it is found that only a few biomarkers are present in the entire gene expression dataset. Hence, if a meta-heuristic-based feature selection algorithm identifies such biomarkers, it would be helpful for the medical experts for proper diagnosis of the disease. An algorithm's stability in feature selection is usually measured by checking the overlap of features selected over multiple independent runs of the meta-heuristic. This work assesses the feature selection stability using the Jaccard Index (JI) between pairs of best-fit solutions over 20 independent runs of the AltWOA over the datasets. That is, 20 best-fit solutions are stored. The pairwise overlap is computed (the percentage of places where both solutions have selected a feature, i.e., '1' in the binary encoded solution). Such a pairwise JI computation of the best-fit solutions yields a matrix of pairwise stability values, which is then averaged over to calculate the stability index for the AltWOA.

Table 5 tabulates the results obtained on computing the stability indices of the various meta-heuristics on the eight microarray datasets used in this study. We can observe that the proposed feature selection model, AltWOA, behaves more consistently and shows its stability which is more profound than the other popular algorithms used for comparison in this study. The difference in the stability indices is especially pronounced on datasets like Alon [104], Sorlie [108] and DLBCL [111].

5. Conclusion and Future Directions

Redundancy is bound to creep in in this data-driven world where many attributes are freely available. Thus, feature selection, a fundamental paradigm of machine learning, is used to optimize storage requirements by eliminating redundancy in data and sifting out only the relevant attributes. Such methods are crucial for gene expression datasets, where thousands of genes are available, but only a few genes are relevant to the disease under study.

The proposed feature selection algorithm, called AltWOA, has been tested on eight high-dimensional DNA microarray datasets. The performances obtained have been compared to several popularly used meta-heuristics found in the literature, revealing the superiority of the proposed method. In this paper, we have developed an improved version of WOA wherein we have embedded the concept of altruism in its population. We aim to enhance the altruism mechanism by using concepts different from our own as future work. Also, the possibility of extending the proposed method to the multi-objective scenario might be intriguing.

In the future, we can also check the viability of the proposed method on various modalities of datasets which may contain both RNA and DNA sequences. There the model may be estimated on either of RNA or DNA sequence dataset, and can be validated on the other dataset. In addition to that a proper benchmarking of the proposed AltWOA can also be done as future extension of present work.

Acknowledgements

The authors would like to thank the Centre for Microprocessor Applications for Training, Education and Research (CMATER) research laboratory of the Computer Science and Engineering Department, Jadavpur University, Kolkata, India, for providing the infrastructural support.

[1] N. Altman, M. Krzywinski, The curse (s) of dimensionality, Nat Methods 15 (6) (2018) 399–400.
[2] P. Luukka, Feature selection using fuzzy entropy measures with similarity classifier, Expert Systems with Applications 38 (4) (2011) 4600–4607.
[3] K. Kira, L. A. Rendell, A practical approach to feature selection, in: Machine learning proceedings 1992, Elsevier, 1992, pp. 249–256.
[4] S. Mirjalili, A. Lewis, The whale optimization algorithm, Advances in engineering software 95 (2016) 51–67.
[5] S. Mirjalili, S. M. Mirjalili, X.-S. Yang, Binary bat algorithm, Neural Computing and Applications 25 (3) (2014) 663–681.
[6] S. Maldonado, J. López, Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for svm classification, Applied Soft Computing 67 (2018) 94–105.
[7] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of machine learning research 3 (Mar) (2003) 1157–1182.
[8] S. Chattopadhyay, R. Kundu, P. K. Singh, S. Mirjalili, R. Sarkar, Pneumonia detection from lung x-ray images using local search aided sine cosine algorithm based deep feature selection method, International Journal of Intelligent Systems (2021).
[9] A. Kaveh, M. I. Ghazaan, Enhanced whale optimization algorithm for sizing optimization of skeletal structures, Mechanics Based design of structures and Machines 45 (3) (2017) 345–362.
[10] D. Oliva, M. Abd El Aziz, A. E. Hassanien, Parameter estimation of photovoltaic cells using an improved chaotic whale optimization algorithm, Applied energy 200 (2017) 141–154.
[11] I. Aljarah, H. Faris, S. Mirjalili, Optimizing connection weights in neural networks using the whale optimization algorithm, Soft Computing 22 (1) (2018) 1–15.
[12] M. Abdel-Basset, G. Manogaran, D. El-Shahat, S. Mirjalili, A hybrid whale optimization algorithm based on local search strategy for the permutation flow shop scheduling problem., Future Gener. Comput. Syst. 85 (2018) 129–145.
[13] D. Gupta, J. Arora, U. Agrawal, A. Khanna, V. H. C. de Albuquerque, Optimized binary bat algorithm for classification of white blood cells, Measurement 143 (2019) 180–190.
[14] M. Kang, J. Kim, J.-M. Kim, Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm, Information Sciences 294 (2015) 423–438.
[15] D. Tripathi, D. R. Edla, V. Kuppili, R. Dharavath, Binary bat algorithm and rbfn based hybrid credit scoring model, Multimedia Tools and Applications 79 (43) (2020) 31889–31912.
[16] V. Basetti, A. K. Chandel, Optimal pmu placement for power system observability using taguchi binary bat algorithm, Measurement 95 (2017) 8–20.

Table 5: Comparison in the stability of feature selection of AltWOA with the popular meta-heuristics on the datasets used in this study. The stability has been calculated by averaging the pair-wise Jaccard Index of the fittest solutions obtained on 20 independent runs of the algorithms.

Dataset	Stability											
	AltWOA	**WOA**	**BBA**	**CS**	**EO**	**GA**	**GSA**	**HS**	**MA**	**PSO**	**RDA**	**SCA**
Alon [104]	**0.3001**	0.2532	0.2696	0.2278	0.2203	0.2061	0.2665	0.2511	0.2726	0.2312	0.2379	0.2145
Christensen [105]	**0.3291**	0.2221	0.2595	0.2360	0.2289	0.2386	0.2237	0.2071	0.2290	0.2584	0.2637	0.2217
Gravier [106]	**0.3476**	0.2316	0.2756	0.2789	0.2753	0.2649	0.2011	0.2045	0.2795	0.2764	0.2490	0.2621
Khan [107]	**0.3313**	0.2153	0.2617	0.2116	0.2352	0.2516	0.2482	0.2610	0.2345	0.2213	0.2773	0.2040
Sorlie [108]	**0.3642**	0.2661	0.2875	0.2010	0.2424	0.2624	0.2399	0.2399	0.2051	0.2058	0.2104	0.2778
Leukemia [109]	**0.3328**	0.1996	0.1985	0.2076	0.2068	0.2004	0.2788	0.2621	0.2689	0.2799	0.2205	0.2314
11_Tumors [110]	**0.4297**	0.3870	0.3495	0.3803	0.3284	0.3399	0.3879	0.3869	0.3426	0.3643	0.3774	0.3069
DLBCL [111]	**0.3586**	0.2166	0.2031	0.2686	0.2587	0.2028	0.2190	0.2489	0.2112	0.2027	0.2524	0.2275

[17] X.-S. Yang, S. Deb, Cuckoo search via lévy flights, in: 2009 World congress on nature & biologically inspired computing (NaBIC), Ieee, 2009, pp. 210–214.

[18] X.-S. Yang, S. Deb, Engineering optimisation by cuckoo search, International Journal of Mathematical Modelling and Numerical Optimisation 1 (4) (2010) 330–343.

[19] R. A. Vazquez, Training spiking neural models using cuckoo search algorithm, in: 2011 IEEE Congress of Evolutionary Computation (CEC), IEEE, 2011, pp. 679–686.

[20] A. R. Yildiz, Cuckoo search algorithm for the selection of optimal machining parameters in milling operations, The International Journal of Advanced Manufacturing Technology 64 (1) (2013) 55–61.

[21] L. H. Tein, R. Ramli, Recent advancements of nurse scheduling models and a potential path, in: Proceedings of the 6thIMT-GT Conference on Mathematics, Statistics and its Applications (ICMSA2010), Citeseer, 2009.

[22] A. Faramarzi, M. Heidarinejad, B. Stephens, S. Mirjalili, Equilibrium optimizer: A novel optimization algorithm, Knowledge-Based Systems 191 (2020) 105190.

[23] M. Abdel-Basset, R. Mohamed, S. Mirjalili, R. K. Chakrabortty, M. J. Ryan, Solar photovoltaic parameter estimation using an improved equilibrium optimizer, Solar Energy 209 (2020) 694–708.

[24] A. Wunnava, M. K. Naik, R. Panda, B. Jena, A. Abraham, A novel interdependence based multilevel thresholding technique using adaptive equilibrium optimizer, Engineering Applications of Artificial Intelligence 94 (2020) 103836.

[25] D. T. Abdul-hamied, A. M. Shaheen, W. A. Salem, W. I. Gabr, R. A. El-sehiemy, Equilibrium optimizer based multi dimensions operation of hybrid ac/dc grids, Alexandria Engineering Journal 59 (6) (2020) 4787–4803.

[26] A. Dey, S. Chattopadhyay, P. K. Singh, A. Ahmadian, M. Ferrara, R. Sarkar, A hybrid meta-heuristic feature selection method using golden ratio and equilibrium optimization algorithms for speech emotion recognition, IEEE Access 8 (2020) 200953–200970.

[27] J. H. Holland, Genetic algorithms, Scientific american 267 (1) (1992) 66–73.

[28] K.-S. Shin, Y.-J. Lee, A genetic algorithm application in bankruptcy prediction modeling, Expert systems with applications 23 (3) (2002) 321–328.

[29] P. R. Srivastava, T.-h. Kim, Application of genetic algorithm in software testing, International Journal of software Engineering and its Applications 3 (4) (2009) 87–96.

[30] Q. Wang, The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, Water resources research 27 (9) (1991) 2467–2471.

[31] A. Norouzi, A. H. Zaim, Genetic algorithm application in optimization of wireless sensor networks, The Scientific World Journal 2014 (2014).

[32] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, Gsa: a gravitational search algorithm, Information sciences 179 (13) (2009) 2232–2248.

[33] B. Shaw, V. Mukherjee, S. Ghoshal, A novel opposition-based gravitational search algorithm for combined economic and emission dispatch problems of power systems, International Journal of Electrical Power & Energy Systems 35 (1) (2012) 21–33.

[34] S. Mondal, A. Bhattacharya, S. H. nee Dey, Multi-objective economic emission load dispatch solution using gravitational search algorithm and considering wind power penetration, International Journal of Electrical Power & Energy Systems 44 (1) (2013) 282–292.

[35] A. Hatamlou, S. Abdullah, H. Nezamabadi-Pour, Application of gravitational search algorithm on data clustering, in: International conference on rough sets and knowledge technology, Springer, 2011, pp. 337–346.

[36] D. Pelusi, R. Mascella, L. Tallini, J. Nayak, B. Naik, A. Abraham, Neural network and fuzzy system for the tuning of gravitational search algorithm parameters, Expert Systems with Applications 102 (2018) 234–244.

[37] Z. W. Geem, J. H. Kim, G. V. Loganathan, A new heuristic optimization algorithm: harmony search, simulation 76 (2) (2001) 60–68.

[38] K. S. Lee, Z. W. Geem, S.-h. Lee, K.-w. Bae, The harmony search heuristic algorithm for discrete structural optimization, Engineering Optimization 37 (7) (2005) 663–684.

[39] M. T. Ayvaz, Identification of groundwater parameter structure using harmony search algorithm, in: Music-inspired harmony search algorithm, Springer, 2009, pp. 129–140.

[40] A. Panchal, Harmony search in therapeutic medical physics, in: Music-inspired Harmony search algorithm, Springer, 2009, pp. 189–203.

[41] E. Yazdi, V. Azizi, A. T. Haghighat, A new biped locomotion involving arms swing based on neural network with harmony search optimizer, in: 2011 IEEE International Conference on Automation and Logistics (ICAL), IEEE, 2011, pp. 18–23.

[42] K. Zervoudakis, S. Tsafarakis, A mayfly optimization algorithm, Computers & Industrial Engineering 145 (2020) 106559.

[43] S. Kadry, V. Rajinikanth, J. Koo, B.-G. Kang, Image multi-level-thresholding with mayfly optimization., International Journal of Electrical & Computer Engineering (2088-8708) 11 (6) (2021).

[44] D. Wei, J. Ji, J. Fang, N. Yousefi, Evaluation and optimization of pem fuel cell-based cchp system based on modified mayfly optimization algorithm, Energy Reports 7 (2021) 7663–7674.

[45] M. H. Hassan, H. Youssef, S. Kamel, F. Jurado, A new application of mayfly optimization algorithm for parameter estimation of single-phase transformer, in: 2021 12th International Renewable Energy Congress (IREC), IEEE, 2021, pp. 1–5.

[46] Z. Liu, P. Jiang, J. Wang, L. Zhang, Ensemble forecasting system for short-term wind speed forecasting based on optimal sub-model selection and multi-objective version of mayfly optimization algorithm, Expert Systems with Applications 177 (2021) 114974.

[47] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of ICNN'95-international conference on neural networks, Vol. 4, IEEE, 1995, pp. 1942–1948.

[48] G.-F. Fan, L.-L. Peng, X. Zhao, W.-C. Hong, Applications of hybrid emd with pso and ga for an svr-based load forecasting model, Energies 10 (11) (2017) 1713.

[49] A. Godio, A. Santilano, On the optimization of electromagnetic geophysical data: Application of the pso algorithm, Journal of Applied Geophysics 148 (2018) 163–174.

[50] M. Rudek, O. C. Junior, T. Greboge, A pso application in skull prosthesis modelling by superellipse, ELCVIA: electronic letters on computer vision and image analysis 12 (2) (2013) 1–12.

[51] Y.-L. Li, W. Shao, L. You, B.-Z. Wang, An improved pso algorithm and its application to uwb antenna design, IEEE Antennas and wireless

propagation letters 12 (2013) 1236–1239.
[52] A. M. Fathollahi-Fard, M. Hajiaghaei-Keshteli, R. Tavakkoli-Moghaddam, Red deer algorithm (rda): a new nature-inspired meta-heuristic, Soft Computing 24 (19) (2020) 14637–14665.
[53] R. A. Zitar, L. Abualigah, Application of red deer algorithm in optimizing complex functions, in: 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2021, pp. 1–6.
[54] B. Balashunmugaraja, T. Ganeshbabu, Privacy preservation of cloud data in business application enabled by multi-objective red deer-bird swarm algorithm, Knowledge-Based Systems 236 (2022) 107748.
[55] G. N. Nguyen, N. H. Le Viet, A. F. S. Devaraj, R. Gobi, K. Shankar, Blockchain enabled energy efficient red deer algorithm based clustering protocol for pervasive wireless sensor networks, Sustainable Computing: Informatics and Systems 28 (2020) 100464.
[56] K. Renuka, R. Muralidharan, Controlling traffic in internet of vehicles using energy aware optimized intelligent transport system using red deer algorithm with new fnn method, Psychology and Education Journal 57 (8) (2020) 959–974.
[57] S. Mirjalili, Sca: a sine cosine algorithm for solving optimization problems, Knowledge-based systems 96 (2016) 120–133.
[58] S. Gupta, K. Deep, A novel hybrid sine cosine algorithm for global optimization and its application to train multilayer perceptrons, Applied Intelligence 50 (4) (2020) 993–1026.
[59] S. Das, A. Bhattacharya, A. K. Chakraborty, Solution of short-term hydrothermal scheduling using sine cosine algorithm, Soft Computing 22 (19) (2018) 6409–6427.
[60] K. S. Reddy, L. K. Panwar, B. Panigrahi, R. Kumar, A new binary variant of sine–cosine algorithm: development and application to solve profit-based unit commitment problem, Arabian Journal for Science and Engineering 43 (8) (2018) 4041–4056.
[61] R. K. Sahu, B. Shaw, J. R. Nayak, Fractional-order pid controller optimized by sca for solar system, in: International Conference on Artificial Intelligence, Smart Grid and Smart City Applications, Springer, 2019, pp. 1–10.
[62] D. H. Wolpert, W. G. Macready, et al., No free lunch theorems for search, Tech. rep., Technical Report SFI-TR-95-02-010, Santa Fe Institute (1995).
[63] M. Khishe, M. R. Mosavi, Chimp optimization algorithm, Expert systems with applications 149 (2020) 113338.
[64] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications, Future generation computer systems 97 (2019) 849–872.
[65] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, S. M. Mirjalili, Salp swarm algorithm: A bio-inspired optimizer for engineering design problems, Advances in Engineering Software 114 (2017) 163–191.
[66] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, W. Al-Atabany, Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems, Applied Intelligence 51 (3) (2021) 1531–1551.
[67] F. Zitouni, S. Harous, R. Maamri, The solar system algorithm: a novel metaheuristic method for global optimization, IEEE Access 9 (2020) 4542–4565.
[68] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, A. H. Gandomi, The arithmetic optimization algorithm, Computer methods in applied mechanics and engineering 376 (2021) 113609.
[69] W. Li, G.-G. Wang, A. H. Gandomi, A survey of learning-based intelligent optimization algorithms, Archives of Computational Methods in Engineering (2021) 1–19.
[70] D. Yazdani, R. Cheng, D. Yazdani, J. Branke, Y. Jin, X. Yao, A survey of evolutionary continuous dynamic optimization over two decades–part b, IEEE Transactions on Evolutionary Computation (2021).
[71] A. Singh, S. Sharma, J. Singh, Nature-inspired algorithms for wireless sensor networks: A comprehensive survey, Computer Science Review 39 (2021) 100342.
[72] N. Khanduja, B. Bhushan, Recent advances and application of metaheuristic algorithms: A survey (2014–2020), Metaheuristic and Evolutionary Computation: Algorithms and Applications (2021) 207–228.
[73] S. Chattopadhyay, A. Dey, P. K. Singh, Z. W. Geem, R. Sarkar, Covid-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer, Diagnostics 11 (2) (2021) 315.
[74] C. Yan, J. Ma, H. Luo, A. Patel, Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets, Chemometrics and Intelligent Laboratory Systems 184 (2019) 102–111.
[75] P. Lopez-Garcia, A. D. Masegosa, E. Osaba, E. Onieva, A. Perallos, Ensemble classification for imbalanced data based on feature space partitioning and hybrid metaheuristics, Applied Intelligence 49 (8) (2019) 2807–2822.
[76] R. Kundu, H. Basak, P. K. Singh, A. Ahmadian, M. Ferrara, R. Sarkar, Fuzzy rank-based fusion of cnn models using gompertz function for screening covid-19 ct-scans, Scientific reports 11 (1) (2021) 1–12.
[77] R. Kundu, P. K. Singh, M. Ferrara, A. Ahmadian, R. Sarkar, Et-net: an ensemble of transfer learning models for prediction of covid-19 infection through chest ct-scan images, Multimedia Tools and Applications (Aug 2021). doi:10.1007/s11042-021-11319-8.
URL https://doi.org/10.1007/s11042-021-11319-8
[78] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, R. Sarkar, Pneumonia detection in chest x-ray images using an ensemble of deep learning models, PloS one 16 (9) (2021) e0256630.
[79] N. A. Al-Thanoon, O. S. Qasim, Z. Y. Algamal, A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics, Chemometrics and Intelligent Laboratory Systems 184 (2019) 142–152.
[80] A. K. Shukla, P. Singh, M. Vardhan, Gene selection for cancer types classification using novel hybrid metaheuristics approach, Swarm and Evolutionary Computation 54 (2020) 100661.
[81] M. M. Mafarja, S. Mirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection, Neurocomputing 260 (2017) 302–312.
[82] G. Kaur, S. Arora, Chaotic whale optimization algorithm, Journal of Computational Design and Engineering 5 (3) (2018) 275–284.
[83] S. Mostafa Bozorgi, S. Yazdani, Iwoa: An improved whale optimization algorithm for optimization problems, Journal of Computational Design and Engineering 6 (3) (2019) 243–259.
[84] F. S. Gharehchopogh, H. Gholizadeh, A comprehensive survey: Whale optimization algorithm and its applications, Swarm and Evolutionary Computation 48 (2019) 1–24.
[85] S. K. Baliarsingh, S. Vipsita, A. H. Gandomi, A. Panda, S. Bakshi, S. Ramasubbareddy, Analysis of high-dimensional genomic data using mapreduce based probabilistic neural network, Computer methods and programs in biomedicine 195 (2020) 105625.
[86] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, M. Lang, Benchmark for filter methods for feature selection in high-dimensional classification data, Computational Statistics & Data Analysis 143 (2020) 106839.
[87] S. K. Baliarsingh, K. Muhammad, S. Bakshi, Sara: A memetic algorithm for high-dimensional biomedical data, Applied Soft Computing 101 (2021) 107009.
[88] M. Alzaqebah, K. Briki, N. Alrefai, S. Brini, S. Jawarneh, M. K. Alsmadi, R. M. A. Mohammad, I. ALmarashdeh, F. A. Alghamdi, N. Aldhafferi, et al., Memory based cuckoo search algorithm for feature selection of gene expression dataset, Informatics in Medicine Unlocked 24 (2021) 100572.
[89] K. GÜÇKIRAN, İ. Cantürk, L. ÖZYILMAZ, Dna microarray gene expression data classification using svm, mlp, and rf with feature selection methods relief and lasso, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi 23 (1) (2019) 126–132.
[90] C. Koushik, A. S. Madhav, R. K. Singh, An efficient approach to microarray data classification using elastic net feature selection, svm and rf, in: Journal of Physics: Conference Series, Vol. 1911, IOP Publishing, 2021, p. 012010.
[91] S. Begum, R. Sarkar, D. Chakraborty, U. Maulik, Identification of biomarker on biological and gene expression data using fuzzy preference based rough set, Journal of Intelligent Systems 30 (1) (2021) 130–141.
[92] M. Ghosh, S. Begum, R. Sarkar, D. Chakraborty, U. Maulik, Recursive memetic algorithm for gene selection in microarray data, Expert Systems with Applications 116 (2019) 172–185.
[93] N. Almugren, H. Alshamlan, A survey on hybrid feature selection methods in microarray gene expression data for cancer classification, IEEE Access 7 (2019) 78533–78548.

Journal Pre-proof

Journal Pre-proof

[94] K. K. Ghosh, S. Begum, A. Sardar, S. Adhikary, M. Ghosh, M. Kumar, R. Sarkar, Theoretical and empirical analysis of filter ranking methods: Experimental study on benchmark dna microarray data, Expert Systems with Applications 169 (2021) 114485.

[95] A. B. Brahim, M. Limam, Ensemble feature selection for high dimensional data: a new method and a comparative study, Advances in Data Analysis and Classification 12 (4) (2018) 937–952.

[96] B. Pes, Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains, Neural Computing and Applications 32 (10) (2020) 5951–5973.

[97] M. Ghosh, S. Adhikary, K. K. Ghosh, A. Sardar, S. Begum, R. Sarkar, Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods, Medical & biological engineering & computing 57 (1) (2019) 159–176.

[98] H. Güney, H. Öztoprak, Microarray-based cancer diagnosis: repeated cross-validation-based ensemble feature selection, Electronics Letters 54 (5) (2018) 272–274.

[99] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, Y. Saeys, Robust biomarker identification for cancer diagnosis with ensemble feature selection methods, Bioinformatics 26 (3) (2010) 392–398.

[100] G. James, D. Witten, T. Hastie, R. Tibshirani, Resampling methods. an introduction to statistical learning, Springer texts in statistics (2013) 175–201.

[101] M. R. Aniba, O. Poch, J. D. Thompson, Issues in bioinformatics benchmarking: the case study of multiple sequence alignment, Nucleic acids research 38 (21) (2010) 7353–7363.

[102] S. Mangul, L. S. Martin, B. L. Hill, A. K.-M. Lam, M. G. Distler, A. Zelikovsky, E. Eskin, J. Flint, Systematic benchmarking of omics computational tools, Nature communications 10 (1) (2019) 1–11.

[103] L. M. Weber, W. Saelens, R. Cannoodt, C. Soneson, A. Hapfelmeier, P. P. Gardner, A.-L. Boulesteix, Y. Saeys, M. D. Robinson, Essential guidelines for computational method benchmarking, Genome biology 20 (1) (2019) 1–12.

[104] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, A. J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proceedings of the National Academy of Sciences 96 (12) (1999) 6745–6750.

[105] B. C. Christensen, E. A. Houseman, C. J. Marsit, S. Zheng, M. R. Wrensch, J. L. Wiemels, H. H. Nelson, M. R. Karagas, J. F. Padbury, R. Bueno, et al., Aging and environmental exposures alter tissue-specific dna methylation dependent upon cpg island context, PLoS genetics 5 (8) (2009) e1000602.

[106] E. Gravier, G. Pierron, A. Vincent-Salomon, N. Gruel, V. Raynal, A. Savignoni, Y. De Rycke, J.-Y. Pierga, C. Lucchesi, F. Reyal, et al., A prognostic dna signature for t1t2 node-negative breast cancer patients, Genes, chromosomes and cancer 49 (12) (2010) 1125–1134.

[107] J. Khan, J. S. Wei, M. Ringner, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, et al., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nature medicine 7 (6) (2001) 673–679.

[108] T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. Van De Rijn, S. S. Jeffrey, et al., Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, Proceedings of the National Academy of Sciences 98 (19) (2001) 10869–10874.

[109] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, science 286 (5439) (1999) 531–537.

[110] A. I. Su, J. B. Welsh, L. M. Sapinoso, S. G. Kern, P. Dimitrov, H. Lapp, P. G. Schultz, S. M. Powell, C. A. Moskaluk, H. F. Frierson, et al., Molecular classification of human carcinomas by use of gene expression signatures, Cancer research 61 (20) (2001) 7388–7393.

[111] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, et al., Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, Nature medicine 8 (1) (2002) 68–74.

[112] J. Krawczuk, T. Łukaszuk, The feature selection bias problem in relation to high-dimensional gene data, Artificial intelligence in medicine 66 (2016) 63–71.

[113] L. I. Kuncheva, A stability index for feature selection., in: Artificial intelligence and applications, 2007, pp. 421–427.

[114] L. Yu, Y. Han, M. E. Berens, Stable gene selection from microarray data via sample weighting, IEEE/ACM Transactions on Computational Biology and Bioinformatics 9 (1) (2011) 262–272.

Journal Pre-proof

Journal Pre-proof

Highlights

- DNA datasets have mainly gained significant attention to the research community.
- They are very high-dimensional, while only a few of those are “bio-markers".
- To solve this problem, it is used the Altruistic Whale Optimization Algorithm
- It effectively filters out only the relevant genes from a large set of attributes.
- Our method reveals its superiority compared to popular and classical techniques in the literature.

Journal Pre-proof

Journal Pre-proof

A DISCLOSURE / CONFLICT OF INTEREST STATEMENT:

• None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.
• It is to specifically state that “No Competing interests are at stake and there is No Conflict of Interest” with other people or organizations that could inappropriately influence or bias the content of the paper.

Journal Pre-proof