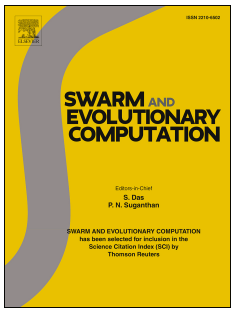

Gene selection for cancer types classification using novel hybrid metaheuristics approach

Alok Kumar Shukla, Pradeep Singh, Manu Vardhan

PII: S2210-6502(19)30283-4

DOI: https://doi.org/10.1016/j.swevo.2020.100661

Reference: SWEVO 100661

To appear in: *Swarm and Evolutionary Computation BASE DATA*

Received Date: 17 April 2019

Revised Date: 10 December 2019

Accepted Date: 2 February 2020

Please cite this article as: A.K. Shukla, P. Singh, M. Vardhan, Gene selection for cancer types classification using novel hybrid metaheuristics approach, *Swarm and Evolutionary Computation BASE DATA* (2020), doi: https://doi.org/10.1016/j.swevo.2020.100661.

Gene selection for cancer types classification using novel hybrid metaheuristics approach

Alok Kumar Shukla[a], Pradeep Singh[a], Manu Vardhan[a]

[a]*Department of Computer Science & Engineering, NIT Raipur, Chhattisgarh, India*

Abstract

With the advancement of microarray technology, gene expression profiling has shown remarkable effort to predict the different types of malignancy and their subtypes. In microarrays, predicting highly discriminative genes is a challenging task and existing hybrid methods fail to deal with efficiently. To mitigate the curse of dimensionality problem and to improve the interpretability of discriminative genes, in this study, we developed a new hybrid wrapper approach which integrates the characteristics of teaching learning-based algorithm (TLBO) and gravitational search algorithm (GSA), called TLBOGSA. A new encoding strategy is also integrated into TLBOGSA to transmute the continuous search space to binary search space and form binary TLBOGSA. In the proposed method, firstly, minimum redundancy maximum relevance (mRMR) feature selection is employed to select relevant genes from the gene expression datasets. Then, wrapper method is applied to select the informative genes from the reduced data produced by mRMR. To improve the search capability during the evolution process, we have incorporated the gravitational search mechanism in the teaching phase. The proposed method uses naive bayes classifier as a fitness function to select the extremely judicious genes which can help to classify cancer accurately. The efficiency of proposed method is tested on ten biological datasets and compared with state-of-art computational intelligence approaches for tumor prediction. Experimental results and statistical analysis demonstrate that proposed method is significantly outperforms existing metaheuristic approaches regarding convergence rate, classification accuracy and optimal number of feature sets. The proposed method reaches above 98 % classification accuracy in six datasets and maximum accuracy is achieved as 99.62 % in DLBCL dataset.

Keywords: Accuracy, Classification, Gene Selection, Naive Bayes, Teaching Learning-based Optimization.

*Corresponding author
**Alok Kumar Shukla
Email address: akshukla.phd2015.cs@nitrr.ac.in (Alok Kumar Shukla)

NOMENCLATURE					
C	Cost Parameter	R^n_M	Gene expression space with M sample and n genes	L_k	Lower Bound with k learner
$M^t_{p,i}$	Passive gravitational mass associated to ih particle at time t	$M^t_{a,j}$	Active gravitational mass associated to jth particle at time t	U_k	Upper Bound with k learner
$R^t_{i,j}$	Euclidian distance between two particles i and j at time t	$F^k_{i,j}$	Gravitational forces from agent j on agent i at a specific time t	G (t)	Gravitational constant at time t
m	Instances/samples	c_1 and c_2	Acceleration parameters	D	Dimension
β	Selected genes	X_p and X_q	Two random learners from X	$M_{i,k}$	Mean of ith learner kth dimension
ζ_i	Support Vectors	μ_p	Mutation Rate	T	Activation function
b	Bias factor	T_E	Estimation time for SVM	T_f	Teaching factor
ρ^n_j	Feature space with j sample and n features	T_F	Time for Updating operations	T_P	True Positive
l^i	Number of labels	T_g	Estimation time in the fitness evaluation	T_N	True Negative
α	Constant Value	ω	Inertia weight	F_P	False Positive
d	Constant parameter	Pm	mutation probability rate	F_N	False Negative
Etime	Execution Time	c	Class /Label	ϑ	Chromosome length
Acc	Accuracy	Z	Minimal redundancy	$nPop$	Population Size
$\aleph$	Maximum number of generation/iterations	V^k_i	The velocity of ith learner's kth dimension	w	Max Dependency
$X_{i,k}$	ith learner kth dimension	$X_{teacher,k}$	Best solution in kth iteration	γ	Classification accuracy
ρ	Rate with which the pheromone assigned	$ace^k_i(t)$	acceleration of i particle with k dimension at time t	Pc	Crossover probability rate

1. Introduction

In the field of bioinformatics, cancer is a remarkably wide group of diseases that covers different characteristics which leads to their different situation categorized by their tissue or organ of origin. Over the past few decades, many experiment-based studies have been conducted to reveal the clue of cancer subtypes [1]. Breakthroughs on high throughput technologies, such as gene expression profiling, provided opportunities for reform treatment by analyzing the endogenous expression levels of thousands of genes in single experiment[2]. A major application of gene expression profiling is cancer classification. In general, the selection and classification of genes is a difficult task from microarray datasets, because the datasets have thousands of the genes and limited number of samples. Generally, in high-dimensional data, irrelevant and redundant genes do not only decrease the training strength but also negatively influence the performance of learning algorithms which is mainly caused by curse of dimensionality [3]. In order to address these problems, researchers have been investigated large number of gene selection methods, many of them derived from the need to analyze microarray data to select the best discriminating gene called biomarker [4]. In classification problem, gene selection is a central application of data reduction to avoid challenges, such as over fitting, high computational burden, and low interpretability of the final model [5][6].

Rapid advances in micro array-based technologies during the last two centuries have opened up new possibilities for life scientists to gather massive data in DNA sequencing, bio imaging, and medical imaging domains. To determine robust gene signature and accurate samples prediction in biological data, many techniques of gene selection have been proposed. Generally, gene selec-

tion methods are assembled into four sets such as filter, wrapper, hybrid, and embedded methods [7]. In general, wrapper and embedded approaches utilize gene selection as a part of training the learning model whereas filter approaches determine the genes independently from a classification model. Filter approaches are known open-loop scheme. It is a straightforward method and observes the features established on the intrinsic behavior of the dataset before the learning tasks[8]. Furthermore, the filter models only depend on on the intrinsic properties of the training data to select genes that have high score value according to the ranking criterion function, they are not influenced by the classifiers [9].

To be more specific, wrapper method use classification algorithm as a fitness evaluation to find the optimal number of the subset of genes through all features and also enhance the classification performance. The results of the wrapper methods generally provides better accuracy, but computational level more expensive than the filter methods [10][11]. Among wrapper-based algorithms [12], gravitational search algorithm (GSA) has been the focus of a great deal of research because of their superior global search capabilities and faster convergence in bioinformatics domains. Moreover, the leading advantages of GSA are less computational effort, and very few number of control parameters compared to other existing evolutionary algorithms (EAs) namely genetic algorithm (GA), ant colony optimization (ACO), differential evolution (DE), simulated annealing (SA), and whale optimization algorithm (WOA) [13][14].

For classifying gene expression profiles, wrapper approaches have applied to select the genes using conventional fitness equation which aim to maximize the performance of classifier only without relatively few feature subsets. Therefore, in this study, we have introduced a new fitness function in order to overcome basic fitness or conventional fitness function limitations. For solving the real-world problems, a new powerful metaheuristic algorithm is investigated, called TLBO in 2011 [15]. Keeping this in mind, TLBO as wrapper approach is used in this paper for getting the near-optimal solutions in the massive search space efficiently. Although, TLBO is a highly effective technique to find the optimum gene subsets, but it have few drawbacks such as premature convergence and slow convergence which encourages us to look for further progress. Thus, a new framework is still required to address the high-dimensional problem from the bioinformatics datasets.

Recent wrapper advancements in bioinformatics have opened up possibilities not only to study biological systems from a comprehensive perspective, but provided unparalleled access to molecular details of active organisms. in available scientific literature, several landmark wrapper FS methods have been investigated for gene selection to cancer/ tumor classification [16] but these techniques still suffer from the absence of correlation between the genes in the search process and also increase the computational burden for finding the best genes. To overcome these shortcomings, researchers have been investigated many hybrid evolutionary algorithms, for example, hybrid of DE and ABC [17], hybrid crow search algorithm[18], TLBOSA[19], in addition, EFS and

AGOA algorithm [20], and fusion of BBHA and BPSO[21].

Most of the reported techniques still suffer from a lot of deficiencies, such as high execution time and stuck in local optima; they do not achieve satisfactory classification results because they do not consider the minimization of the size of the selected target gene, and they need a maximum fitness evaluation value and parameters for the adjustment. Therefore, this study developed a new hybrid wrapper algorithm to overcome the limitations of the conventional algorithms and to select the useful genes for accurate cancer subtypes prediction. As far as we know, this is the first time that TLBO and GSA work together and form a hybrid wrapper model which resolved the limitations of individual algorithms. In the proposed strategy, GSA is considered as a local search, and hybridize with TLBO to overcome the issue as stuck in local optima by motivating the exploration of the search. Moreover, this paper introduces new encoding, new fitness function, and new updating scheme for solving the gene selection problems.

The proposed algorithm has some potential advantages such as increases the interpretability of the dataset, reduces the computational complexity, and control the premature and stagnation problems. In addition, it has the ability to search globally and to handle the high-volume data with acceptable solutions within a reasonable time. The identification of discriminating genes are fundamental and practical interest. Through mRMR method in proposed method searches the high ranked genes that are passed into the wrapper (TLBOGSA) approach, it can be benefit from the examination of genes in biology and medicine to recent findings in cancer research or suggest new ways to explore. The proposed method is combined with naive bayes classifier which helps for finding the most significant genes that can attempt to accurate cancer prediction. This procedure continues until a reasonable accuracy is achieved with small gene subsets. The main highlights of this paper as follow:

- This paper introduced a new hybrid wrapper algorithm, called TLBOGSA by combination of TLBO and GSA methods. It can improve the convergence rate and balance the trade-off between exploration and exploitation capabilities.
- A new encoding scheme, new updating mechanism for learners and new fitness function is also introduced for better cancer classification.
- The proposed method uses a more compatible fitness function as naive bayes to estimate the classification performance.
- The proposed method selects highly discriminative gene subsets from various microarray datasets and is less prone to over-fitting than using all gene space. It is capable to achieve comparable classification accuracy in comparing with well-known wrapper methods.

Through the above analysis, we can be observed that several hybrid techniques have built cancer diagnosis frameworks, but these frameworks have some flaws that are worth improving.

The main objective of this study is to develop a diagnostic modal based on a few possible genes which are capable to detect cancer subtypes accurately. In order to obtain more explicable genes and overcome the deficiencies of the existing investigated methods, in this study, a new hybrid gene selection method by combination of TLBO and GSA technique is proposed. By identifying a small subset of genes on which to base diagnosis, we not only enhance the classification performance but also obtain potentially significant information about the nature of the disease and the genetic mechanisms responsible for it. Furthermore, this research opts the small subset of gene information from microarray data that maximizes the accuracy to make a diagnosis more likely to be displayed in a clinic. To the best of our knowledge, we are one of the first group to develop a new wrapper method called TLBOGSA in biological datasets for gene selection.

The remainder of the arrangement is planned as follows. Section 2 and Section 3 begin with a brief introduction to feature subset selection, followed by a characteristic of TLBO, GSA and other methods, which is essential to wrapper approach. Section 4, we describe a hybrid algorithm (TLBOGSA) with NB classifier and it merits. Section 5, we present the experimental results and discussions on the gene expression data sets. Section 6 elaborate the conclusion.

2. Related Work

From the machine learning point of view, one of the most desirable tasks is to divide genes into different groups according to the gene relevancy. However, it is difficult to detect the gene relevancy for complex biological structure, in which metaheuristic methods have provided a new insight. The diversity of metaheuristic algorithm is applied to other optimization problems such as large-scale global optimization functions, combinatorial optimization, and bioinformatics problems [22]. An explicit disadvantage associated with existing approach is that the algorithms cannot guarantee the optimal feature subset because the most compact and informative feature subset is unknown in real-world applications. To avoid filtering out a few critical genes highly related to sample classes and simplify the process of gene selection, in this study, we have analyzed and discussed new evolutionary algorithms with limitations of the previously studied method. Among existing evolutionary algorithms, teaching-learning-based-optimization (TLBO) plays an important role for solving discrete and binary feature selection problems because it is easy to implement and have less amount of control parameters[23].

The original TLBO was proposed for solving the continuous optimization problems, while this study emphases on gene selection process in microarray gene expression classification which is consider as combination problem with high dimensionality. Basically, it is a human-based approach that works on the teaching and learning phenomena, where teachers affect the quality of results of learners. One of TLBO variant was investigated by [24] called neighbor teaching-learning-based-optimization (NTLBO), which is applied for improving the performance of basic

TLBO when they solved the global optimization problems. Investigated algorithm focused on teaching-learning phenomena to learn the knowledge from the classmate, and also each student signified a possible solution of the optimization problems. The experimental results convey that TLBO outperformed the existing metaheuristic methods when proposed method solved the unconstrained eight test functions. Recently, the performance of TLBO algorithm was estimated on 25 numerical test suites against other metaheuristic algorithms such as PSO and GA. The results of experimental study shown that TLBO outperforms the PSO and GA algorithm regarding convergence solution [19].

Shahbeig et al.[25] have introduced combination of the teaching learning-based optimization and diffused particle swarm optimization (PSO) applied to find the least subset of genes intricate in breast cancer with the most significant amount of precision. In addition to this, proposed approach have applied to solve the multi-objective FS and clustering problems[26]. In the available literature, numerous modification in TLBO has been done which shown the better performance as compared to original TLBO in terms of accuracy, optimal number of genes, and stability. Similar to other EAs, TLBO method is also stuck in local optima when they have resolved the complex numerical problems. A new physics-based algorithm has introduced in the recent time called gravitational search algorithm (GSA) that were introduced by Rashedi, in 2009 [27]. According to [12], in machine learning domains, a new FS method called FSS-MGSA has investigated for feature subset selection which was successfully applied to high dimensional search space. As compared to other evolutionary techniques, FSS-MGSA method has some merits like a simple structure, dominant optimal performance, rapid convergence speed, and easy to understand. In 2012, a new modal investigated by applying position-based learning GSA [28] and Immune gravitation optimization algorithm (IGOA)[29] to enhance the convergence speed of GSA. The mentioned study incorporated the features of anti-body diversity and vaccination into the GSA algorithm.

For cancer diagnosis, a new computational gene selection model for microarray data classification through adaptive hypergraph embedded dictionary learning has proposed [30]. Specifically, a dictionary was learned from the feature space of original high dimensional microarray data and this learned dictionary was used to represent original genes with a reconstruction coefficient matrix. With a massive influx of multi-modality data, the role of data analytics in health informatics has grown rapidly in the last decade [31]. From [32] author demonstrated a new strategy that applied deep learning to an ensemble approach that incorporated multiple different machine learning models. Author supplied informative gene data selected by differential gene expression analysis to five different classification models. Then, a deep learning method was employed to ensemble the outputs of the five classifiers. The proposed deep learning-based multi-model ensemble method was tested on three public RNA- seq data sets of three kinds of cancers, Lung Adenocarcinoma, Stomach Adenocarcinoma and Breast Invasive Carcinoma.

To address inaccessibility of optimal feature subset size, Wang et al. [33] have presented a

weighted general group lasso model to opt cancerous genes types. The proposed method was investigated based on weighted gene network analysis and also a new method for calculating gene and group weights is introduced in terms of joint mutual information in 2018. To decrease the high computational cost associated with feature selection algorithm, Wang et al.[34] have presented the hybrid GSA-PSO for improving the solution quality obtained by GSA for solving function optimization problems in 2017. To select the optimal subsets of genes from ensemble approach, Emmanuella et al. [35] have presented a ensemble gene selection approach using particle swarm optimization, ant colony optimization, and genetic algorithm methods. Similarly, in this study, author proposed novel consensus gene selection criteria for partial least squares-based gene microarray analysis. By quantifying the extent of consistency and distinctiveness of the differential gene expressions across different double cross validation or randomizations in terms of occurrence and randomization p-values, proposed criteria were able to identify a more comprehensive genes associated with the underlying disease. A distributed GPU implementation has proposed to accelerate the gene selection problem and about 8-11 times speed up has achieved based on the microarray datasets considered [36].

In the available literature, several population-based optimization algorithms namely ABC, DE, GA, PSO, and GSA have been applied to solve classification problems [37]. One of the nature-inspired algorithm as binary ABC method utilizes the benefits of bees to resolve the gene selection problem. However, binary DE can be applied for solving the discrete issues using a probability estimation operator[38]. However, binary GA solves real-world optimization problems such as combinatorial and numerical problems[39]. The operational structure of binary PSO and GSA is similar to solve the mathematical global optimization problems[40]. In 2014, a hybrid modal PSOGSA were examined by Wang et al.[41] that combined the assets of both particle swarm optimization and GSA for solving the global optimization problem. Similarly to PSOGSA [42], social intellectual and individual thoughtful of PSO were integrated to GSA for solving a continuous problem. In addition, Nasir and Tokhi[43] have presented a new hybrid GS model as spiral-dynamic bacteria-chemotaxis algorithms which were applied to address the global optimization problem and their application to control of a flexible manipulator system.

By applying computational intelligence methods in medical data helps acquire better controlling, faster performance and higher level of accuracy in detection. In [44], author introduced the state-of-the-art of computational intelligence approaches in medical data and to categorize the existing computational intelligence techniques, used in medical fields with the help of single and hybrid method. In addition, the techniques and methodologies, their limitations and performances were presented. From [1] a new bacterial colony optimization method with multi-dimensional population, called BCO-MDP was presented for feature selection for the purpose of classification. To address the combinational problem associated with feature selection, population with multiple dimensionality was represented using different feature subsets sizes. Author [45] have

proposed framework (C-HMOSHSSA) for gene selection using multi-objective spotted hyena optimizer (MOSHO) and salp swarm algorithm (SSA). The real-life optimization problems with more than one objective usually face the challenge to maintain convergence and diversity. Salp swarm algorithm maintains diversity but suffers from the overhead of maintaining the necessary information. On the other hand, the calculation of MOSHO requires low computational efforts is used for maintaining the necessary information. Therefore, proposed algorithm was hybrid algorithm that utilized the features of both SSA and MOSHO to facilitate its exploration and exploitation capability.

3. Background Details

3.1. Feature Selection

The predictions of classification techniques are deeply influenced by the quality of the input gene space. The primary goal of feature selection (FS) is to select the meaningful and relevant features from the original feature space with the minimum redundancy and the maximum discriminating ability. Due to redundant features and curse of dimensionality, learning engine takes significant amount of time and performance of the model decreases. In other words, Feature selection is one of the essential pre-processing tools in bioinformatics to solve this problem and enhance the classification performance [46]. By reducing the irrelevant and meaningless features, FS process diminishes the data dimensionality, reduce the quantity of data needed for any learning algorithm, reduce the execution time, simplify the structure and improve the performance of the learned classifiers[7]. It has verified from the literature that "classifications done with feature subsets given as an output of FS have higher prediction accuracy than classifications carried out without FS" [47].

Metaheuristic techniques have recently received much attention from the feature selection community as they are well-known for their global search ability/potential. Finding an optimal gene subset for a given problem, several metaheuristic algorithms such as GA, GSA, TLBO, DE, and PSO have been discovered for optimizing of FS problems [48][49]. In recent decades, TLBO and GSA are two well-known metaheuristic methods in the field of computational intelligence. In recent studies, there are no results on an combination of TLBO and GSA on high dimensional datasets which works as wrapper. This paper puts ahead a gene selection method based on hybrid of TLBO and GSA consider for ten microarray gene expression data classification and examines its performance.

3.2. Teaching Learning-based Optimization

Teaching Learning-based Optimization (TLBO) algorithm has been recently proposed in the literature as novel population oriented metaheuristic algorithm. The improper tuning of the control parameters enhances the computational burden or produces the local optimal result. Among

existing metaheuristic, TLBO is a simple yet global optimization method that has been successfully employed in some unconstrained and constrained non-linear complex numerical problems, including some design optimization problems with considerable success in early 2012 [15]. TLBO algorithm is based on the effect of influence of a teacher on the output of the learners in a class. In this algorithm, the population maintained as group of learners or class of learners. The primary aim of this algorithm is to find the best learner through cooperation and the sharing of information between the students. The working behavior of TLBO algorithm depends on the extremely sophisticated learners so that they can produce better results regarding marks or grades. The group of the learners also called as a class. They learn through neighbor learning phenomena. The whole population is represented as a vector describing in Equation 1.

$$X_{i,k} = \begin{bmatrix} x_{1,1} & x_{1,2} \cdots & x_{1,k} \\ \vdots & \ddots & \vdots \\ x_{\mathrm{i},1} & x_{\mathrm{i},2} \cdots & x_{\mathrm{i},k} \end{bmatrix} \tag{1}$$

Where i represents the number of populations, k is the dimension of the problem, and $X_{i,k}$ is the position of the ith learner in the kth dimension. The learner X is randomly initialized within search space. The evolution of $X_{i,k}$ is randomly generated by the following Equation 2.

$$X_{i,k} = L_k + r_1 * (U_k - L_k) \tag{2}$$

Where i = 1, 2, 3,...,nPop, k=1,2,3...,D, r_1 signifies random variable, L_k signifies the lower bound, and U_k signifies the upper bound value. The simulation of classical learning process to all learners is categories into two significant phases of TLBO algorithm: Teacher Phase and Learner Phase. The optimum learner is obtained from the teacher phase through the knowledge learning from the neighbors and the learner phase find the best learner through the interaction between different learner groups. The basic approach of the TLBO method is explained in the following sections:

3.2.1. Teacher phase

In this phase, the teacher upgrades knowledge and initiate messages through the class, therefore, improve the grade mean of the whole class. The teacher is an amount of gaining optimum result obtained so far from the optimization problems. As a good teacher updates knowledge through the learners knowledge. A teacher can increase the mean result of class to a certain value which depends on the capability of the whole class. Let $M_{i,k} = (1/nPop)(\sum X_{i,k})$ be the mean value of the particular subject where k = 1, 2,..., D. The updating equation of process as described in Equation 3.

$$X_{i,k}^{new} = X_{i,k}^{old} + r_2 * (X_{teacher,k} - T_f * M_{i,k}) \tag{3}$$

Here, $X_{teacher,k}$ is the best learner of adopted the population at kth iteration of the algorithm, r_2 random number in the range [0,1]; the value of T_f is considered as 1 according to [15].

Algorithm 1 Basic teaching learning based optimization

Set the size of population as nPop, dimension as D, and the maximum number of iterations as t_{max}, L_k , and U_k;
for i= 1 to nPop **do**
evaluate the $X_i^{old} = L_k + rand^* (U_k - L_k)$
end for
f = objective function (X_i^{old})
Estimate $X_{mean} = \sum X_i^{old} / nPop$
$X_{teacher}$ = the best learner from the whole class;
while (t $\leq t_{max}$) **do**
t=t+1;
$Diff_{Vec}$ = ($X_{teacher}$ - T_f * X_{mean})
for i= 1 to nPop **do**
for k= 1 to D **do**
$X_{i,k}^{new}$ = ($X_{i,k}^{old}$) + rand * $Diff_{Vec}$(k)
f = objective function ($X_{i,k}^{new}$)
end for
if f ($X_{i,k}^{new}$) < f($X_{i,k}^{old}$) **then**
$X_{i,k}^{old} = X_{i,k}^{new}$
end if
end for
for i= 1 to nPop **do**
Randomly select the other two learner X_p and X_q
for k= 1 to D **do**
if f (X_p) > f(X_q) **then**
$X_{i,k}^{new}$ = $X_{i,k}^{old}$ + rand * (X_p-X_q)
else
$X_{i,k}^{new}$ = $X_{i,k}^{old}$ + rand * (X_q - X_p)
end if
end for
end for
Store the highest fitness value ;
end while

3.2.2. Learner Phase

The second phase of TLBO algorithm is learner phase that increases the knowledge of learners by the two different ways: one way is through input from the teacher, and the next method is through mutual interaction between them. The goal of each learner is randomly interacted with peer learners and enhance communication grade. To select the ith learner as X_p and another random learner is X_q(p≠q) through the mutual interaction with learners. The updating equation of the ith learner X_p in the learning phase can described in Equation 4 and Equation 5.

if $f((X_p) < f(X_q))$

$$X_{i,k}^{new} = X_{i,k}^{old} + r * (X_p - Xq) \tag{4}$$

else

$$X_{i,k}^{new} = X_{i,k}^{old} + r * (X_q - Xp) \tag{5}$$

end if

Where r represents a random number between [0,1]; $f(X_p)$ and $f(X_q)$ are the best solution of the learners X_p and X_q .

3.3. Gravitational Search Algorithm

Over the past few years, various metaheuristic methods have been developed. Many of these optimization methods are inspired by swarm behaviors in nature. In 2009, a new optimization algorithm called gravitational search algorithm (GSA) was introduced [27] based on the law of gravity and mass interactions. In GSA, the agents are a collection of masses which interact with each other based on the Newtonian gravity and the laws of motion. Further, each agent is attracted by all the other agents. Newton's universal law of gravity ascertains the gravitational force acting on each factor and the speed of each factor is calculated according to Newton's law of movement. After applying gravitational force, the agents can investigate the search space to search for the optimal solutions. As compared to other evolutionary methods, GSA has numerous attractive features, such as easy to implement, reliable performance, and smaller number of tuning constraints.

Suppose a system with N agents; the algorithm jumps by the randomly generated value of individuals in the whole search space. The gravitational forces (F) from agent j on agent i at a specific time t is defined as the duration of learning by the Equation 6.

$$F_{i,j}^{d}(t) = G(t) \frac{M_{pi}^{t} * M_{aj}^{t}}{R_{pi}^{t} + e} \left(x_j^d(t) + x_i^d(t) \right) \tag{6}$$

Where, M_{aj}^{t} represent the active gravitational mass related to agent j, M_{pi}^{t} represent the passive gravitational mass related to agent i, G (t) shows the gravitational constant at time t, e is a small

constant, and $R_{i,j}^{t}$ represent the Euclidian distance between two agents i and j. The formula of $G(t)$ is estimated according to Equation 7.

$$G(t) = G_0 * exp(-alpha * \frac{iter}{\aleph}) \tag{7}$$

Where alpha value taken directly from [50] as 20, G_0 set to 100. *iter* shows the current iteration, and $\aleph$ present the maximum number of generations. The total force on particle i is calculated according to Equation 8.

$$F_i^d(t) = \sum_{j=1, j\neq i}^{N} rand_j F_{i,j}^d(t) \tag{8}$$

Where, $rand_j$ shows the random value lies between 0 and 1. As reported by law of motion concept, the acceleration of an agent is relative to the resultant force and inverse of its mass, so the acceleration of all agents can measured as Equation 9.

$$ace_i^d(t) = \frac{F_i^d(t)}{M_{i,i}^t} \tag{9}$$

Where t represents the specific time, $M_{i,i}^t$ represent the mass of object i, and $rand_j$ represent the random value in the interval [0, 1]. Firstly, all masses are initialized with random numbers in GSA. Each mass represents the as candidate solution. The gravitational constant, total forces, and accelerations are calculated Equation 7 to Equation 9. According to Equation 10 and Equation 11, we can update the gravitational and inertial masses, respectively.

$$m_i(t) = \frac{f_i(t) - worse(t))}{(best(t) - worse(t))}, where \;\; i = 1, 2, ..., nPop \tag{10}$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{nPop}(m_j(t))}, where \;\; j = 1, 2, ..., nPop \tag{11}$$

Where $f_i(t)$ signify the fitness value of the agent i at time t, and, worst (t) and best (t) are defined according to Equation 12 and Equation 13:

$$best(t) = min_{j\in(1,...,nPop)} f_j(t), where \;\; j = 1, 2, ..., nPop \tag{12}$$

$$worse(t) = max_{j\in(1,...,nPop)} f_j(t), where \;\; j = 1, 2, ..., nPop \tag{13}$$

Furthermore, the next velocity of an agent is considered as a fraction of its current velocity added to its acceleration. Therefore, its position and its velocity could be calculated according to Equation 14 and Equation 15.

$$v_i^d(t+1) = rand_i * v_i^d(t) + ace_i^d(t) \tag{14}$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1) \tag{15}$$

where $rand_i$ is a uniform random variable in the interval [0,1]. We use this random number to give a randomized characteristic to the search.

3.4. Deep Learning Techniques

Deep learning, which is especially challenging in handling high dimensional dataset, has achieved great impact in the field of health informatics [32]. At the heart of theoretical foundations of deep learning is inspired by classical neural network (NN) literature. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction but different to more traditional use of NNs. Up to now several deep learning models have been investigated, the most typical deep learning models include, Deep Neural Networks (DNNs), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), which are widely used in gene expression profiling. From [51], researcher introduced a comprehensive up-to-date evaluation on the research employing deep learning in health informatics, providing a critical analysis of the relative worth, and likely pitfalls of the technique as well as its future outlook. This study mainly focused on key applications of deep learning in the fields of bioinformatics, medical informatics, and public health.

3.4.1. Deep Neural Network

Deep neural networks (DNNs) refer to end-to-end mappings by stacking a large number of filters learned from large records. It is currently widely used for many applications including computer vision, bioinformatics, and string matching [52]. While DNNs present state-of-the-art accuracy on many evolutionary tasks, it comes at the cost of high computational complexity. Accordingly, techniques that enable efficient processing of DNNs to improve efficiency and throughput without sacrificing application accuracy or increasing hardware cost are critical to the wide deployment of DNNs in evolutionary system. The basic structure of DNN consists of an input layer, multiple hidden layers, and an output layer. Once input data are given to the DNNs, output values are computed sequentially along the layers of the network. In each layer, the input vector comprising the output values of each unit in the layer below is multiplied by the weight vector for each unit in the current layer to acquire the weighted sum. Then, nonlinear function such as sigmoid, hyperbolic tangent, and rectified linear unit, is applied to the weighted sum to calculate the output values of the layer. The computation in each layer transforms the representations in the layer below into slightly more abstract representations. Based on the layers types used in DNNs and the comparable learning method, DNNs is widely classified in auto-encoder multilayer perceptron, deep belief network.

3.4.2. Convolutional Neural Network

In general, so far investigated exclusive DNNs cannot scale well with multi-space input that has locally correlated data due to train large number of nodes and parameters. To resolve the previous NNs problem, scientist introduced a new deep learning method called Convolutional Neural Networks (CNNs) [32] to analyze biomedical data. The architecture of CNNs is analogous to that of the connectivity pattern of neurons in the human brain and is directly inspired by the visual cortex of the brain. Individual neurons respond to stimuli only in a restricted region of the visual field. A collection of such fields overlap to cover the entire visual area. In spite of the CNNs advancement in recent years, many applications are still untouched to handle real-world data without manual human labels. CNN is able to successfully capture the spatial and temporal dependencies in multiple data types through the application of relevant filters. The basic cognition of CNNs consists of convolution, nonlinear, and pooling layers. To fully commute the two-dimensional structure of an input data, local connections and shared weight in the network are used, instead of traditional fully connected networks. This process results in very few parameters to learn which makes the network faster and easier to train. Nonlinear layers enhance the nonlinear properties of feature maps. In each pooling layer, maximum or average sub-sampling of non-overlapping regions in feature maps is performed. This non-overlapping sub-sampling enables CNNs to handle slightly different but semantically similar features and accordingly combined local features to identify more complex features.

3.4.3. Recurrent Neural Network

Recurrent Neural Networks (RNNs) are popular models, are designed to commute sequential information, have basic structure with cyclic connection that shown great promise in many evolutionary tasks [53]. It is type of Neural Network where the output from previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other. The main and most important feature of RNN is hidden layer which remembers some information about a sequence. Since input data are processed sequentially, recurrent computation is performed in the hidden units where cyclic connection exists. Therefore, ancient information is implicitly stored in the hidden layers called state vectors, and output for the current input is computed considering all previous inputs using these state vectors. Although the importance of deep learning is increasing and several advances in its research are touching great heights, but still there have few downsides or challenges to be tackled to develop it.

Still, researchers have been found the open research challenge in wrapper and hybrid techniques such as scalability, parameter control, exploration and exploitation dilemma, evaluation measure, and representation to solve the feature selection problems. Unfortunately, none of the existing metaheuristic algorithms are entirely perfect due to uncertainty encompassing to predict the tumor subtypes with high accuracy at low computational burden. So, the need to develop pro-

ficient wrappers has raised constantly. Therefore, this study has developed a new hybrid wrapper algorithm, is used for selection of candidate biomarkers.

4. Proposed approach for gene selection

From machine learning standpoint, having too many genes always leads to over-fitting and curse of dimensionality and a negative influence on classification. Indeed, the presence of redundant features may adversely affect the classification accuracy, as they can add more noise than useful information. Several hybrid algorithms for cancer classification that employ evolutionary algorithm as wrapper have been proposed in the literature which can be able to overcome the mentioned challenge and select highly potential features from gene expression data for enhancing the predictive performance for cancer classification. Keeping this in mind, this study introduced a new wrapper algorithm by hybridization of TLBO and GSA in such a way that they complement to each other for finding the better quality of solution regarding classification accuracy and minimize the volume of irrelevant features. In addition, new encoding scheme and fitness function are also investigated to estimate the classification performance.

The proposed method has the following aspects such as per-processing, optimization, and classification. As a preliminary consideration to processing, the original data are per-processed by the filter as mRMR method. Each gene can have estimated according to information theory, and the ranking of genes are ordered based on the used filter method and then create a reduced dataset. The most advantageous aspect of mRMR method is applied to filter the noisy and redundant genes and, in turn, reduce the computational cost of the used wrapper.

4.1. Pre-processing Phase

To effectively filter out the highly redundant and irrelevant genes, usually, filter-based feature selection approaches are utilized for tumor classification [54]. We observed that the minimization of max-dependency and maximal relevance on the gene sets is hard to understand. To handle this kind of problem, we used a more efficient method as minimum Redundancy Maximum Relevance (mRMR) for gene selection in this study as filter. In the proposed method, mRMR [36] is worked before the wrapper method to generate high quality of gene solutions and thus permit to the search process. For gene variables, the mutual information (MI) of two variables X and c is defined based on their entropy of X and c is updated according to Equation 16.

$$I\left(c;X\right)=H\left(c\right)-H\left(\frac{c}{\mathrm{X}}\right) \tag{16}$$

Where, $H\left(c\right)$ and $H\left(\frac{c}{x}\right)$ is the entropy and conditional entropy between the class and variable. By using mutual information concept, researchers have designed a gene selection with the aim is

to choose a subsets S with N genes with maximum dependency on the target class c; so-called Max Dependency, is formulated as Equation 17.

Algorithm 2 Standard gravitational search algorithm

Begin
Set the initial value of gravitational constant (G_0), ϵ, α ,t_{max}, nPop;
Set the initial iteration t=0;
for i= 1 to nPop **do**
Generate an initial population $X_i(t)$ randomly, where $[X_i(t) = x_i^1(t), x_i^2(t), x_i^3(t), ..., x_i^n(t)]$;
end for
Evaluate the fitness value as Z = fitness ($X_i(t)$, ξ) for each agent
Assign the best, worse agent in the $X_i(t)$ population
X^* = the best search learner;
while (t $\leq t_{max}$) **do**
t=t+1;
Calculate the gravitational constant G(t) according to Equation 7;
Update best (t) and worse (t) according to Equation 12 and Equation 13;
for i = 1 to nPop **do**
Estimate M(i) according to Equation 11;
end for
for i= 1 to nPop **do**
Calculate all gravitational force $f_i^d(t)$ by Equation 6;
Calculate acceleration $ace_i^d(t)$ by Equation 9;
Calculate position X_i by Equation 14;
Update velocity $v_i^d(t)$ according to Equation 15;
end for
end while
Until the predefined number of the run is attained
Return the best solution ;

$$max w\,(X,c) = I\,(c; x_1, x_2, \ldots, x_N) = H\,(c) - H\left(\frac{c}{x_1, x_2, \ldots, x_N}\right) \tag{17}$$

According to Equation 17, we can estimate the dependency among features; it can produce large value. The relationship between redundancies between features is expressed as Equation 18

to Equation 19.

$$\min Z\left(X, c\right) = 1/\mid s^2 \mid \sum_{x_j \in s} I\left(x_j; x_k\right) \tag{18}$$

$$Max\varphi\left(w, Z\right) = w - Z \tag{19}$$

The integration of Equation 18 and Equation 19 is known as "minimal-redundancy-maximal-relevance" (mRMR) which describes in Equation 20.

$$j_{mRMR}\left(\varphi\right) = I\left(c; X\right) - 1/\mid s^2 \mid \sum_{x_j \in s} I\left(x_j; x_k\right) \tag{20}$$

Where, x_j is selected subset of gene S and x_k is original genes set.

4.2. Gene-optimization Process

The filter method like mRMR can effectively diminish the dimensionality of the gene size and provide the set G1($\subseteq$ G) of meaningful genes in the form of reduce dataset. The reduced dataset D contains N gene expression vector each of dimensionality m (m $\leq$ M). Now, we analyses the effectiveness of genes classification model to further optimize the dimensionality of G1 gene subgroup, using hybrid wrapper approach, for which NB model achieves the maximum classification accuracy. It is used to estimate the predictive fitness of the optimal subset of genes is performed to reduce the risk of over-processing.

By making a reduction of the data by mRMR method used in the first stage, it gives us the possibility that the information obtained as a result of this pre-processing is information without noise and redundant information, but it is still impossible to gather relevant statistics from this result. In order to reduce the number of selected genes and take the most informative genes with highest classification accuracy and low computational efforts, in this paper, we proposed new hybrid model, called TLBOGSA.

The developed hybrid framework utilizes an efficient strategy for combining two metaheuristic approaches namely TLBO and GSA in order to address the problem of gene selection to discriminating the samples from high dimensional data sets and estimate the classification accuracy. Algorithm 2 can be used to show the Pseudo-code of proposed algorithm. To balance between the exploitation and exploration, at the beginning of BTLBOGSA algorithm, all populations are initialized randomly. Each learner is considered as a candidate solution. After initialization, the gravitational force, the gravitational constant, and the effects between the agents are calculated using Equation 6 to Equation 8. Next, the individual accelerations are defined as Equation 9. In each iteration, the best solution should be updated. After calculating the accelerations and updating the best solution, the new solution for each population is calculated using Equation 21 and Equation 22. The process of updating the learners that meet a final criterion. The representation

of each learner in this proposal is in the form of the binary string of n bits and encoded by using Equation 24. The fitness value of each learner is calculated from Equation 26. The whole process is repeated ten times and selected different genes from the evolutionary algorithm in each dataset. Algorithm 3 can also be used to show the pseudo code of proposed algorithm.

4.2.1. Binary TLBOGSA

Initially, TLBOGSA has developed for solving the complex global optimization problems[50]. In the original TLBOGSA technique, the individuals continuously move around the search space because of having position vectors with the real continuous domain. To require the search agents to move in a binary search space, we possess an improved velocity updating scheme. For this reason, we have presented a binary variant of TLBOGSA called BTLBOGSA. According to[55], a transfer function is additionally required to alter the position of the agent with the possibility of its velocity. Transfer functions can transform discrete values into binary values. A proper transfer function should become capable of providing a substantial probability of adjusting the position for massive significant benefits on the velocity. On the teacher and learner phases, we all expressed the value of velocity as Equation 21 to Equation 22.

$$V_i^{k+1} = r_1 * \left(X_{teacher,k} - T_f * M_{i,k} \right) + r_2 * \left(ace_i^k (t) - T_f * M_{i,k} \right) \tag{21}$$

$$V_i^{k+1} = r * \left(X_p - X_q \right) \tag{22}$$

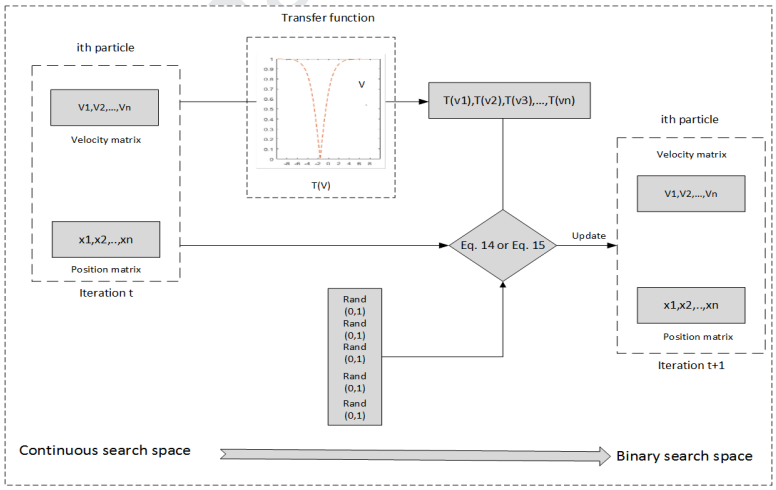

Figure 1: The process of mapping a continuous search space to a discrete search space.

4.2.2. New Encoding Scheme

A binary vector shows the position in proposed method, and the velocity is still floating-point vector, however; velocity is employed to look for the possibility to change from zero to one or one to zero when one of the positions of learner can be updated. A transfer function uses to transform the probability of swapping a position vectors aspects from zero to one and vice versa can be seen in Figure 1. In the available literature, the most popular activation function is a sigmoid function[56] which is used to normalize the velocity in the between 0 to 1. Usually, the disadvantage relates to sigmoid function is that there is no difference among a significant value in Vi,d in the positive and negative direction and then it shows the fact that the greater movement is required based on the previous position. To overcome this problem, we have suggested new transfer functions to the component of velocity. Finally, the proposed V-shaped transfer function is highlighted in Equation 23.

$$T(V_i^k) = \exp\left(\left|\left(V_i^{k+1} - a\right) / \ (1+b)\right|\right) - 1/\exp\left(\left|\left(V_i^{k+1} - a\right) / \ (1+b)\right|\right) + 1 \tag{23}$$

Where a and b are pre-defined constant value its remain static in the whole search process. After calculating the probabilities, the learner updates their positions based on the rules present in Equation 24.

$$X_i^k = \begin{cases} 1, if \ rand < \ T\left(V_i^k\right) \\ 0, \quad otherwise \end{cases} \tag{24}$$

Algorithm 3 BTLBOGSA for feature selection

```
Run=1;
Begin
Initialize the parameters nPop, D, ξ, t_max , t=0;
for i= 1 to nPop do
   X_i = initialize(P);
end for
Z = fitness (X_i, ξ)
X* = the best search learner;
while (t ≤ t_max) do
   t=t+1;
   Calculate the gravitational constant G(t) according to Equation 7;
   Update best (t) and worse (t) according to Equation 12 and Equation 13;
   for i = 1 to nPop do
      Estimate M(i) according to Equation 11;
   end for
   for i= 1 to nPop do
      Calculate all gravitational force f_i^t by Equation 6;
      Calculate acceleration a_i^t by Equation 9;
      Calculate position X_i by Equation 4;
      Calculate probability of changing position vector according to Equation 24;
      Update velocity vector according to Equation 21 and Equation 22;
   end for
   for i=1 to nPop do
      Calculate the fitness of all updated learners according to Equation 26;
   end for
   Update learner X_i and solution Z
end while
Store the optimal feature subset with the highest fitness value ;
run=run+1;
Until the predefined number of the run is attained
```

The primary idea of the hybrid BTLBOGSA is to use GSA as a local optimizer to develop a highly effective gene selection algorithm with better balance among exploration and exploitation including a better convergence rate. Steps of picking the features aided by the TLBO algorithm are revealed in Algorithm 1 and also the GSA algorithm granted in Algorithm 2. These algorithms form the purpose of the proposed algorithm. The overall work of the proposed method is

described according to Algorithm 3.

4.2.3. Basic Fitness Function

To maximize the classification performance regarding accuracy, this paper has employed wrapper (BTLBOGSA) approach which plays a significant role in gene selection. The supreme goal of BTLBOGSA is to select a subset of features to attain better classification accuracy compared to the use of all available features. The basic fitness function (see equation 25) is to maximize the classification accuracy of the algorithm achieved through the selected genes during the evolutionary process.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{25}$$

Where T_p , T_N, F_P, and F_N represent as True Positive, True Negative, False Positive, and False Negative.

4.2.4. New Fitness Function

The optimal feature subsets nominated by merely fitness function may still have potential redundancy since the essential functions do not aim to reduce the number of features. We are hypothesized that the comparable classification accuracy may obtain from a subset of smaller characteristics. To meet this challenge, a new fitness function is investigated with the objectives of increasing the classification accuracy and minimizing the number of genes. As we can see in Equation 26, the newly present fitness function is:

$$fitness\,(x) = \alpha^* \frac{\beta}{\vartheta} + (1 - \alpha)\,^*\gamma \tag{26}$$

Here, fitness (x) denotes the competence of the feature subset represented by x; γ represents the classification accuracy using the Naive Bayes (NB) classifier, ϑ is the measurement of chromosome length, and β is an upper bound of selected feature from the candidate solutions, and α represent the constant value lie between 0 to 1.

Table 1: Dataset Description

No.	Dataset	Instances	Genes	Classes	No.	Dataset	Instances	Genes	Classes
1	Leukaemia_1	72	5327	3	6	SBRCT	83	2308	4
2	Colon-cancer	62	2000	2	7	Lung-cancer	203	12600	5
3	DLBCL	77	5469	2	8	Brain Tumor_1	90	5920	5
4	Leukaemia_2	72	11225	3	9	9_Tumors	60	5726	9
5	Prostate_Tumor	102	10509	2	10	11_Tumors	174	12533	11

Table 2: Parameters setting of existing metaheuristic methods and proposed method

S. No.	Parameters	Value	S. No.	Parameters	Value
1.	Population Size	20	6	ω	0.7
2.	Number of generations	100	7	ρ	0.7
3.	Number of runs	10	8	Pc	0.7
4.	Chromosome length	50	9	Pm	0.4
5.	Performance	Accuracy	10	c_1 and c_2	2

5. Experimental Result and Discussion

In this section, we evaluated proposed approach and other wrapper methods on ten gene expression datasets namely Colon, DLBCL, SBRCT, Prostate-tumour, 9-Tumors, 11-Tumors, Brain Tumor-1, Leukaemia-1, Leukaemia-2 and Lung-cancer. More detailed information about the datasets, we can found in www.gems-system.org. Table 1 summarizes some basic information about the datasets, including the number of features and samples. Out of the ten datasets, three datasets belong to binary-class and seven datasets belong to multi-class. At first, the training group was then divided into two sub-data groups, including training and testing samples sub-data by using tenfold cross-validation. The training sub-data were used only for constructing model and evaluating individuals during the evolutionary process, while the test sub-data were used to assess the final solutions which were in the repository. Immediately after partitioning, the dataset was shrunk concerning predetermined number of top 50 genes by mRMR method as described in Pre-processing Phase, in the current study.

5.1. Parameter Setting

Our execution platform was carried out on the MATLAB 2016a mathematical development environment. The empirical evaluation was performed in this work on desktop with an Intel Core i7 processor, 2.4 GHz and 6 GB of RAM. Table 2 comprises the parameter values for TLBO, GSA, BTLBOGSA, and other nature-inspired algorithms. The values were chosen based on the results of several preliminary runs. We have tested the performance of diverse robust and straightforward classifiers using tenfold cross-validation and to find out which classifier gives the higher performance than remaining methods on top dimensional biological datasets.

5.2. Evaluation Criteria

Classification is a characteristic of learning algorithms. Given a set of instances represented by features and corresponding labels, classification involves learning a model to correctly predict the class membership of each instance. The performance of proposed algorithm is estimated by the NB classifier with the help Equation 26 which works as a fitness function. The proposed method is

also evaluated by four measures that are, Sensitivity, Specificity, Matthews Correlation Coefficient (MCC), and F-measure. These performance measures are defined according to Equation 27 to Equation 30.

$$Sensitivity(Se) = \frac{T_P}{T_P + F_N} \tag{27}$$

$$Specificity(Sp) = \frac{T_N}{T_N + F_P} \tag{28}$$

$$F - measure(Fmes) = \frac{2 * T_P}{2 * T_P + F_N + F_P} \tag{29}$$

$$MCC = \frac{(T_N * T_P - F_N * F_P)}{\sqrt{(T_P + F_P) + (T_P + F_N) + (T_N + F_P) + (T_N + F_N)}} \tag{30}$$

Here, T_P, T_N, F_P, and F_N are True positive, True negative, False positive, and False negative in the independent datasets. Based on confusion matrix[57], we assessed the performance of the proposed method and contend gene selection.

5.3. Experimental Results and Analysis

In this experimental section, the proposed method is validated by comparing it against a series of algorithms on numerous frequently used datasets into three parts: Firstly, applied the four classifiers and choose the best one which is act as a fitness function in filter and wrapper methods. Then, apply the selected classifier as a fitness function in filter-based methods and our approach. The fitness value is evaluated by using the selected classifier NB according to Equation 26. Finally, selected fitness function used to form new fitness function for wrapper algorithms.

In order to regulate which classification algorithm is more robust and has higher performance than the others to be used as fitness function of our algorithm and current optimization algorithms, we have compared four different classifiers such as k-NN, NB, SVM, and DT on ten biological datasets. This numerical analysis is assessing and shown in Table 3. In the next experimental study, our method is compared with filter-based methods, is shown in Table 4. For all datasets, the number of populations for each nature-inspired algorithm is set to twenty. In high dimensional data such as microarrays, most of the genes are meaningless and not suitable for model construction. To select of top most significant genes from DNA microarrays by removing a massive number of irrelevant, redundant, and noisy genes can provide better performance. Therefore, we have chosen the top 50 genes by mRMR method. The parameters of the nature-inspired algorithm are adjusted according to[58].

To prove the strength of our method, and to provide a better comparison of the results obtained with the other algorithms, we have run all algorithms as ten times. When the BTLBOGSA

is applied to the datasets, optimal gene subsets are selected resulting in increased prediction accuracy. To show the effectiveness of BTLBOGSA algorithm, when we used the Friedman test to calculate the power level of classification as statistically significant. It describes the experimental benefits that are obtained by merely our algorithm. From Table 3, we can see that NB classifier gives the good classification performance in terms of accuracy on almost all datasets. The maximum classification accuracy is 98.57 % in SBRCT dataset with the help of the NB classifier, and the lowest classification accuracy is achieved through DT classifier as 73.12 % in 9-tumors. Therefore, we choose a robust classifier as Naive Bayes fitness function in our and another population based algorithm.

Table 3: Average classification performance by using four classifiers on ten biological datasets

Dataset	**Classifier**					**Dataset**	**Classifier**				
	Measures	NB	SVM	kNN	DT		Measures	NB	SVM	kNN	DT
Colon	Acc	94.78	93.97	94.01	93.13	Lung	Acc	90.72	88.43	86.38	84.36
	Se	93.25	91.36	93.45	91.63		Se	88.36	87.36	85.75	82.36
	Sp	91.56	92.21	91.25	92.58		Sp	89.65	88.65	84.96	81.96
	Fmes	88.98	92.86	94.97	93.14		Fmes	91.56	89.39	85.32	83.07
	MCC	89.52	93.52	94.26	95.63		MCC	92.35	93.56	94.78	95.37
Leukemia 1	Acc	94.97	93.47	92.87	91.05	Brain Tumor 1	Acc	86.87	80.57	82.98	81.06
	Se	94.11	92.98	91.35	90.36		Se	85.69	78.35	80.52	82.25
	Sp	93.36	92.65	91.8	91.87		Sp	88.07	77.95	79.35	79.38
	Fmes	93.01	93.01	91.54	92.56		Fmes	86.95	79.36	78.99	79.99
	MCC	94.28	95.89	96.71	96.85		MCC	90.52	91.74	90.25	91.12
Leukemia 2	Acc	92.98	90.58	91.42	89.62	11 Tumors	Acc	89.01	88.34	84.37	86.33
	Se	91.56	88.36	90.36	85.36		Se	88.36	89.36	83.65	85.36
	Sp	92.65	89.02	91.32	87.36		Sp	90.36	90.32	82.89	83.87
	Fmes	93.87	87.65	90.37	88.65		Fmes	90.85	89.35	80.78	82.65
	MCC	94.52	93.52	96.25	89.63		MCC	90.32	88.25	89.36	88.25
DLBCL	Acc	96.07	88.01	94.88	93.51	9 Tumors	Acc	75.34	77.45	78.98	73.12
	Se	95.36	87.32	95.36	92.68		Se	74.63	78.03	77.36	74.36
	Sp	94.87	85.35	93.89	91.95		Sp	73.85	75.85	76.89	71.25
	Fmes	93.65	93.87	92.67	94.57		Fmes	74.89	76.89	77.02	74.36
	MCC	92.25	93.84	94.63	95.12		MCC	75.26	77.96	76.25	78.23
SBRCT	Acc	98.57	96.52	96.22	97.04	Prostate cancer	Acc	93.21	91.32	90.65	92.01
	Se	97.56	93.86	95.63	95.68		Se	92.36	90.55	89.65	92.88
	Sp	95.85	94.86	94.25	96.03		Sp	90.35	89.65	91.03	93.01
	Fmes	98.51	92.86	96.03	97.58		Fmes	91.58	88.39	92.85	92.87
	MCC	97.25	98.26	97.53	96.74		MCC	90.36	91.54	89.57	90.86

To study another test is to TLBO with individual GSA and BTLBOGSA wants to assess the result concerning hybridizing GSA criteria while using TLBO as shown in Table 5. The last try things out is going to demonstrate effectiveness concerning the hybrid BTLBOGSA way compared with other gene collection methods, this includes ACO, GA, SA, PSO, and DE see in Table 6.

Table 4: Average classification accuracy with Top 50 genes from the ten biological datasets

Methods	Leukaemia_1	Colon	DLBCL	SBRCT	Lung	Leukaemia_2	Brain Tumor_1	11_Tumors	9_Tumors	Prostate cancer
CMIM	81.01	80.42	92.54	92.99	90.52	90.67	83.65	84.36	68.95	88.64
JMI	70.54	73.62	78.66	90.54	91.67	91.53	81.36	88.95	64.98	87.34
mRMR	83.23	81.63	88.52	93.48	91.36	92.52	83.65	85.95	69.85	86.85
DISR	76.52	79.33	80.01	92.37	91.55	90.94	81.65	80.11	70.92	81.97
Relief-F	73.99	82.57	80.67	83.54	85.66	80.74	84.03	81.96	71.35	88.16
Proposed	96.01	95.31	98.37	99.54	95.01	96.01	88.54	90.12	77.12	95.36

In this study, we use the five existing filter methods such as CMIM, JMI, mRMR, DISR, and Relief-F to select the top ranked meaningful genes from the ten biological datasets. The comparative evaluation of the proposed method and filters was performed for 50 gene subgroups selected for all datasets. From Table 4, the result shows that the proposed method is superior to other methods of gene selection mentioned on 50 gene subgroups. The proposed method attains the maximum accuracy as 99.54 % in SRBCT dataset in comparison to other methods. We designed experiments on all gene datasets using ten times. Compared to previous reported study, it should be noted that using mRMR gene selection method from Meyer et al.[58] obtained 78.31 % accuracy with NB on the SBRCT data; the proposed approach produce the result as 99.54 % accuracy with only 50 selected genes on the same dataset. The highest efficiency is accomplished by the proposed method as 96.01 % in Leukaemia-1, 95.31 % in Colon, 98.37 % in DLBCL, 99.54 % in SRBCT, 95.01 % in Lung cancer, 96.01 % in Leukaemia-2, 88.54 % in Brain Tumor-1, 90.12 % in 11-Tumors, 77.12 % in 9-Tumors, and 95.36 % in Prostate cancer dataset. This table showed the accuracy results NB classifier except for proposed method which used new fitness function; and selected the number of genes by CMIM, JMI, mRMR, DISR, Relief-F, and proposed method of all datasets.

5.4. Comparative Results Analysis of the wrapper method

Recently, several studies have focused on enhancing the feature subset quality through wrapper approach. Unfortunately, this phenomena is still theoretical or empirical unexplored to study the hybrid wrapper approaches in high dimensional datasets. Therefore, potential feature selection techniques are needed to select optimal feature subsets which have different quality. However, existing wrapper algorithms have been selected gene subsets but they have often different collections of local optima identified within the space of the feature subsets, and their performance is unreliable and unpredictable with high computational time. This motivates us to originate a hybrid wrapper method which can help to identify the perilous genes and classify a disease correctly with less computational time. The results for comparisons are illustrated in Table 5, we have reported the classification accuracy (# Acc) with standard deviation (STD), few features (# Feat), and average execution time (Etime) for 100 generations each of the proposed and the TLBO and GSA algorithms for ten microarray datasets. These results indicate that the TLBOGSA method has the ability of selecting those predictive genes highly related to sample classes.

Table 5: Comparison of the proposed method with another state of arts regarding best classification accuracy with STD and the optimal number of selected features

Dataset	Performance	Methods			Dataset	Performance	Methods		
		Proposed	TLBO	GSA			Proposed	TLBO	GSA
Leukaemia_1	#Acc ± STD	94.15 ±0.07	87.35 ±0.09	84.36 ±1.07	Lung	#Acc ± STD	99.61 ±0.02	90.25 ±0.53	87.96 ±1.15
	#Feat	16	19	21		#Feat	13	18	24
	Etime	25.56	36.35	42.65		Etime	32.85	34.71	38.98
Colon	#Acc ± STD	98.87 ±0.10	95.68 ±0.52	92.47 ±1.34	Brain tumor_1	#Acc ± STD	96.92 ±0.83	88.36 ±1.07	84.65 ±1.79
	#Feat	16	18	25		#Feat	15	21	23
	Etime	10.52	15.68	21.75		Etime	27.85	31.85	34.87
DLBCL	#Acc ± STD	99.62 ±0.04	97.26 ±0.86	94.36 ±1.52	11_tumor	#Acc ± STD	93.04 ±1.75	89.36 ±2.53	84.20 ±2.87
	#Feat	17	20	21		#Feat	13	14	15
	Etime	25.98	31.56	37.85		Etime	40.85	57.03	59.84
SBRCT	#Acc ± STD	99.17 ±0.06	94.36 ±0.97	90.65 ±1.16	9_tumor	#Acc ± STD	70.88 ±2.87	72.36 ±2.52	68.35 ±2.81
	#Feat	11	21	25		#Feat	12	18	27
	Etime	11.58	23.80	27.81		Etime	32.87	37.52	40.86
Leukaemia_2	#Acc ± STD	98.84 ±0.84	91.35 ±0.83	88.35 ±1.67	Prostate cancer	#Acc ± STD	98.42 ±0.67	90.02 ±1.27	87.96 ±2.53
	#Feat	12	18	20		#Feat	07	17	21
	Etime	30.87	33.95	37.85		Etime	30.52	39.24	44.58

In this experimental study, we have calculated classification performance based on new fitness function with the help of NB classifier on ten gene expression datasets. From Table 5, it can be observed that the like TLBO method has like to similar accuracy than the proposed method for 9-tumor dataset. In addition, we can observed that the accuracy of the classification is not very interesting, especially for the 9-Tumors data set. The highest accuracy in Table 5 as 99.62 % in DLBCL with 17 genes.

However, the experimental results on gene datasets are shown in Figure 2 regarding classification performance such as an average number of features and average fitness value. Figure 2 (a) compares the results obtained by proposed, TLBO, and GSA on ten microarray datasets. As seen in Figure 2 (b), the proposed method provides better results than TLBO and GSA regarding a minimum number of selected genes for nine microarray datasets. The diagram shows that the proposed method performed significantly better than both TLBO and GSA algorithm regarding fitness value.

Table 6: The comparative performance of five nature inspired methods and the proposed method

Dataset	Performance	GA	PSO	DE	ACO	SA	Proposed
Leukaemia_2	#Acc ± STD	94.15 ±2.51	87.23 ±2.89	87.35 ±2.57	86.32 ±2.74	84.36 ±2.63	98.84 ±0.84
	#Feat	16	21	19	25	21	12
Colon	#Acc ± STD	92.17 ±0.96	93.64 ±0.87	91.68 ±1.63	90.99 ±1.52	88.47 ±2.13	98.87 ±1.10
	#Feat	21	25	17	18	27	16
DLBCL	#Acc ± STD	97.52 ±0.36	95.42 ±1.52	96.21 ±1.61	94.76 ±2.03	90.52 ±2.83	99.62 ±0.04
	#Feat	20	24	23	25	21	17
SBRCT	#Acc ± STD	95.17 ±1.52	94.22 ±2.07	94.36 ±2.21	90.27 ±2.51	92.65 ±2.85	99.17 ±0.06
	#Feat	17	19	23	21	25	11
Leukaemia_1	#Acc ± STD	97.84 ±0.04	98.23 ±1.81	91.35 ±1.87	90.87 ±1.56	88.35 ±2.87	94.15 ±0.07
	#Feat	19	21	18	27	29	16
Lung	#Acc ± STD	95.61 ±0.53	97.05 ±0.87	91.25 ±0.92	88.46 ±1.31	87.96 ±2.02	99.61 ±0.02
	#Feat	16	19	18	23	24	13
Brain tumor_1	#Acc ± STD	91.92 ±0.86	90.41 ±1.03	87.36 ±1.53	89.56 ±2.52	82.65 ±2.31	96.92 ±0.83
	#Feat	17	18	21	23	23	15
11_tumor	#Acc ± STD	91.04 ±0.87	88.06 ±1.53	83.36 ±1.93	90.65 ±2.53	84.2 ±2.53	93.04 ±1.75
	#Feat	18	20	14	19	26	13
9_tumor	#Acc ± STD	80.88 ±2.53	81.36 ±2.53	72.36 ±2.57	71.23 ±2.63	68.35 ±2.97	70.88 ±2.87
	#Feat	27	22	18	24	27	12
Prostate cancer	#Acc ± STD	90.42 ±0.53	91.91 ±1.53	90.02 ±1.82	86.35 ±2.03	87.96 ±2.34	98.42 ±0.67
	#Feat	14	19	17	20	24	07

Table 7: Comparison of the proposed method with other methods in each dataset regarding accuracy and number of features. The symbol "*" means that no information is available

Methods	Colon	SBRCT	Lung	DLBCL	Leukemia_1	Leukemia_2	Brain_Tumor 1	9_Tumors	11_Tumors	Prostate
GANN[59]	*	59.3 (19)	*	*	47.4 (21)	*	*	*	*	*
BFO[60]	*	97.50(35)	93.11(39)	98.99(8)	96.19(23)	*	90.37(25)	61.11(29)	72.31(39)	97.42(29)
DRF0-CGS[61]	90(10)	*	98.66(17)	94.67 (11)	91.18 (13)	94.12	*	*	*	97.06
IWSSr[62]	84.0(11)	92.3(13)	*	93.6(10)	97.1(7)	97.3(6)	*	*	*	94.3(10)
PSO dICA[63]	94.73(20)	*	97.95 (25)	94.73 (30)	97 (72)	83.33	*	*	*	96.77
BDF[64]	97.46	*	99.14	89.44	95.81	*	*	*	*	79.5
BBHA-RF[60]	91.41	*	*	*	98.61	*	*	*	*	*
BPSO_TS[65]	*	*	99.52	*	*	*	95.89	81.63	97.35	95.45
BDE-X Rankf[66]	75.0 (3)	*	98.0(3)	92.9 (3)	82.4(7)	*	*	*	*	*
SE1DCNN[67]	84.90	*	*	*	57.87	*	*	*	*	*
SAE[68]	66.67	*	*	*	33.71	*	*	*	*	*
DLFCC[69]	*	*	83.00	*	88.00	*	*	*	*	88.00
Proposed method	98.87(16)	99.17(11)	99.61(13)	99.62(17)	94.15(16)	98.84(12)	96.90(15)	70.88(12)	93.04(13)	98.42(7)

5.5. Evaluation of Proposed method with Past literature

The classification performance of the hybrid method is estimated in two phases. In the first phase (mRMR), the best subset data of the gene are selected. In the second phase, class classification label that uses the information sources of the data set of the gene. Come to the Table 7, which is presented below for the economic analysis and the small subgroups of genes on biomedicine. The first column shows the works reported in the previous literature and compared with the proposed method and the remaining column showed percentages at the accuracy and number of genes (in brackets) obtained for each series of genetic data.

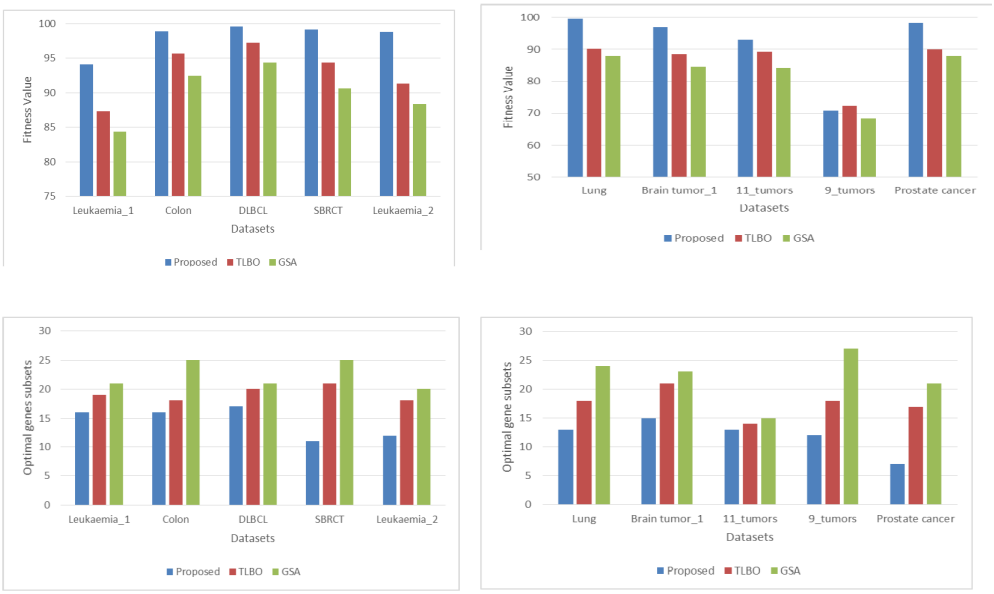

Figure 2: Diagram of a) fitness value and b) optimal subsets for TLBO, GSA, and proposed approach in ten microarray datasets.

Figure 3 shows that the convergence values of BTLBOGSA increase vividly after a few numbers of iterations of almost all the datasets. This leaning indicates that BTLBOGSA is suitable to select a tiny number of genes from high-dimensional data to increase classification performance. The condition of the proposed individual convergence speed that should always be positive real numbers started in the initialization method, and the new rule for updating individual positions irritate the early convergence of BTLBOGSA.

The other investigation is to review the hybrid algorithm employing a new fitness function in Table 6. Table 6 shows the maximum classification accuracy and best gene subset (# Feat) from wrapper-based approaches (i.e., GA, PSO, DE, SA, and ACO) and BTLBOGSA gene selection methods on the gene datasets over ten repeated times. The classification accuracy as 99.62 % in DLBCL dataset with 17 genes is the highest achieved by our proposed method. The best results among all gene selection methods have been highlighted and marked in bold type. As we can see in Table 6, the most famous nature-inspired metaheuristic algorithm as GA selects the number of genes is to 14 in a Prostate dataset with classification accuracy of 98.42 %. In addition, the genes corresponding to GA algorithm is 21 with 92.17 % accuracy in Colon dataset. The classification accuracy as 99.61 % in Lung cancer dataset with 13 genes is the second highest achieved by our proposed method. The nature-inspired algorithm produces the third largest classification accuracy

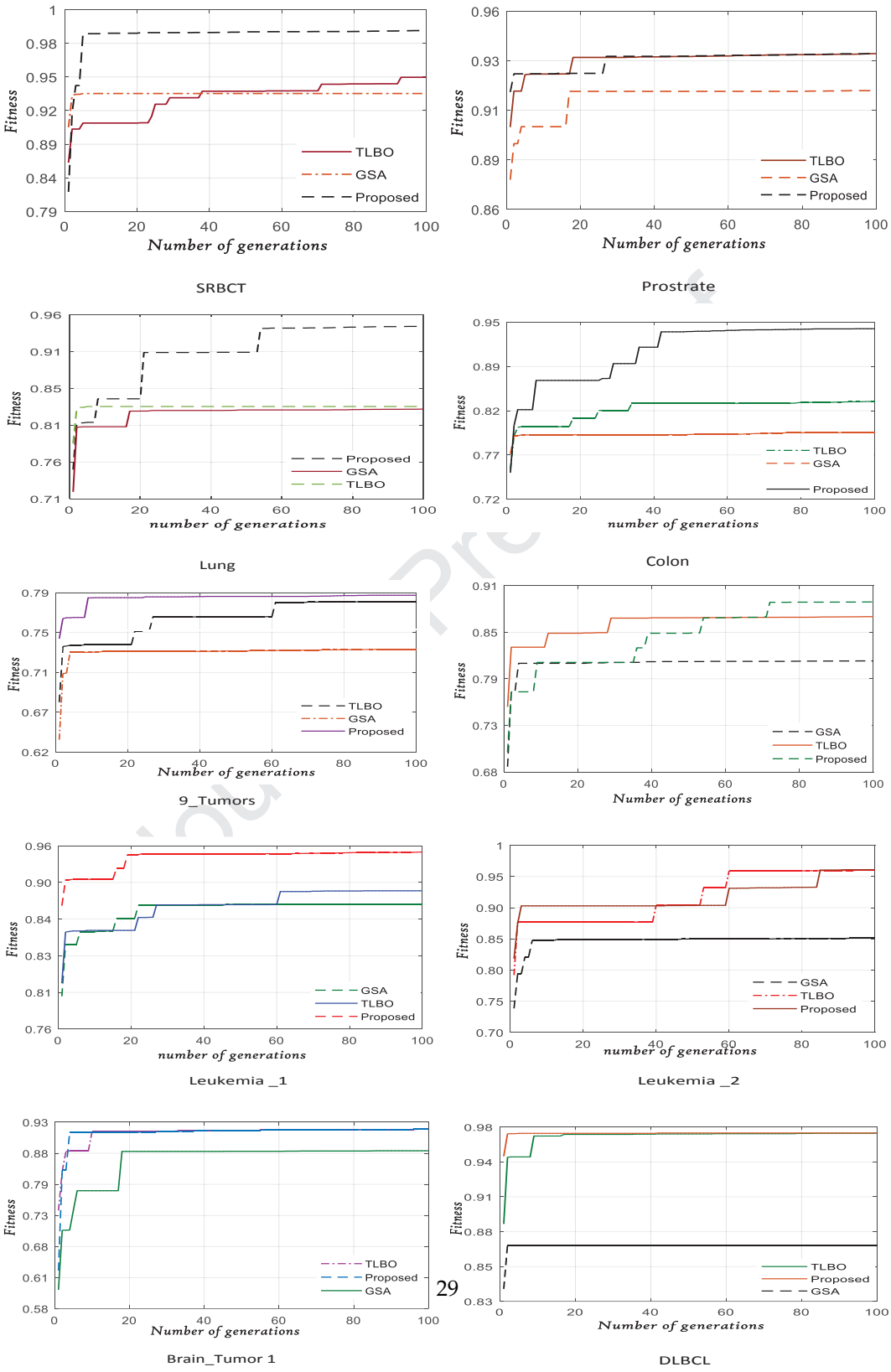

Figure 3: Comparative convergence graphs of proposed method, TLBO and GSA on classification accuracy in ten microarray datasets.

as PSO is 98.23 % with 21 genes in the Leukaemia-1 dataset. In ACO technique, the maximum and minimum classification accuracy is 94.76 % and 71.23 % in DLBCL and 9-tumors dataset. Furthermore, DE algorithm has gained the 96.21 % classification accuracy from the DLBCL dataset with 23 genes.

Table 8: Summary of statistical results.

Algorithm	Ranking
GA	2.5
PSO	3.2
DE	4.75
ACO	5.9
SA	7.3
TLBO	3.85
GSA	6.8
Proposed	1.7

Table 9: Post Hoc comparison Table for α = 0.05

i	Algorithm	z = (R0-Ri) /SE	p	Holm Hochberg	Li
7	SA	5.112077	0.000001	0.007143	0.028147
6	GSA	4.655642	0.000003	0.008333	0.028147
5	ACO	3.834058	0.000126	0.01	0.028147
4	DE	2.784256	0.005365	0.0125	0.028147
3	TLBO	1.962672	0.049684	0.016667	0.028147
2	PSO	1.369306	0.170904	0.025	0.028147
1	GA	0.730297	0.465209	0.05	0.05

5.6. Statistical Results

In general, one of the most frequently used non-parametric statistical method as a Friedman test[70] is applied to rank the algorithm performance. Whether aims to find any remarkable difference exists between the results of different algorithms. It is based on the null hypothesis that there is no dissimilarity in the presentation of algorithms. The best performing algorithm gets the lowest rank while the worst performing algorithm receives the highest rank (see Figure 4) according to following aspects:

1. On observed accuracy (Acc) value for each algorithm and dataset pair.
2. For each gene dataset, rank values from 1 (best result) to 8 (worst result).

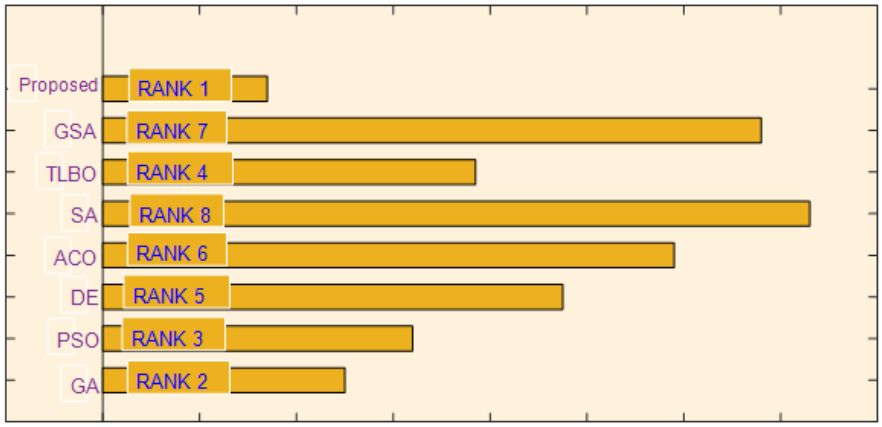

Figure 4: Ranking of existing algorithms and proposed method by Friedman test.

3. For each method, average the ranks obtained in all gene datasets to achieve the final rank.

The tests are applied for the accuracy (Acc) obtained from gene data sets. The average rank obtained by each algorithm on all datasets is evaluated for responsible according to Friedman's statistic which is shown in Table 8.

The computed Friedman statistic is 44.758333 on the reported results, and the p-value computed by test was 0.0054 at 5% level of significance. Therefore, the null hypothesis is rejected, signifying the differences among the performances in the compared techniques. The proposed model was further validated by the posthoc method for pairwise comparisons between the metaheuristic methods using Li and Holm's procedure. P-values obtained in by applying post hoc methods over the results of Friedman procedure as shown in Table 9. Hochberg's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.0125. Li's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.028147.

The experimental results validated that NB is the best classification algorithm and the proposed wrapper based FS approach outperforms the performances of wrappers, and filters in terms of accuracy, MCC, sensitivity, specificity, and the number of selected optimized features. Furthermore, if the execution time is taken into account, proposed approach performs much faster than single TLBO and GSA. TLBOGSA only needs a single parameter for configuring the model and is simple to understand. In the future, the proposed method could be applied in other areas such as pattern recognition and anomaly detection. Also, the improvement in TLBOGSA by applying a new initialization strategy could be treated as a research subject in the future.

6. Conclusion

Recently, biologists and life scientists are attracted in identifying the significant genes related to cancer, which plays a crucial role in the cancer diagnosis and treatment. However, it has been a challenging task to attain those essential genes in performing cancer classification from microarray data due to the high dimensionality of data with limited number of samples. Although the existing wrapper methods are capable of identifying informative genes from the high dimensional datasets, but it has some issues such as expensive computational cost and premature convergence. In order to retain good balance between global search and local search and to identify susceptibility genes of composite diseases, we proposed a new hybrid wrapper modal for gene selection to speed up the learning process, called BTLBOGSA which integrated the characteristics of TLBO and GSA algorithm. In addition, this paper introduced a new encoding scheme and fitness function for finding the relevant genes and achieving higher classification accuracy from DNA microarray datasets. The proposed approach is not only improved the classification performance but also reduced the computational time. Experiments on ten biological datasets demonstrate that the proposed method has excellent performance in terms of both the classification accuracy and small subset of genes and outperforms other existing filters and wrapper methods. The proposed method reaches above 98 % classification accuracy in six datasets. Thus, it can be applied for DNA microarray analysis.

[1] H. Wang, L. Tan, B. Niu, Feature selection for classification of microarray gene expression cancers using bacterial colony optimization with multi-dimensional population, Swarm and Evolutionary Computation 48 (2019) 172–181.

[2] R. Katuwal, P. N. Suganthan, L. Zhang, An ensemble of decision trees with random vector functional link networks for multi-class classification, Applied Soft Computing 70 (2018) 1146–1153.

[3] A. Saha, S. Das, Clustering of fuzzy data and simultaneous feature selection: A model selection approach, Fuzzy Sets and Systems 340 (2018) 1–37.

[4] H. M. Zawbaa, E. Emary, C. Grosan, V. Snasel, Large-dimensionality small-instance set feature selection: A hybrid bio-inspired heuristic approach, Swarm and Evolutionary Computation 42 (2018) 29–42.

[5] A. K. Shukla, P. Singh, M. Vardhan, A hybrid framework for optimal feature subset selection, Journal of Intelligent & Fuzzy Systems 36 (2019) 2247–2259.

[6] S. Paul, S. Das, Simultaneous feature selection and weighting–an evolutionary multi-objective optimization approach, Pattern Recognition Letters 65 (2015) 51–59.

[7] J. C. Ang, A. Mirzal, H. Haron, H. N. A. Hamed, Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection, IEEE/ACM transactions on computational biology and bioinformatics 13 (2016) 971–989.

[8] J. Wang, J.-M. Wei, Z. Yang, S.-Q. Wang, Feature selection by maximizing independent classification information, IEEE Transactions on Knowledge and Data Engineering 29 (2017) 828–841.

[9] H. Kim, S.-M. Choi, S. Park, Gseh: A novel approach to select prostate cancer-associated genes using gene expression heterogeneity, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 15 (2018) 129–146.

[10] G. Wu, R. Mallipeddi, P. N. Suganthan, Ensemble strategies for population-based optimization algorithms–a survey, Swarm and evolutionary computation 44 (2019) 695–711.

[11] A. K. Shukla, Identification of cancerous gene groups from microarray data by employing adaptive genetic and support vector machine technique, Computational Intelligence (2019).

[12] X. Han, X. Chang, L. Quan, X. Xiong, J. Li, Z. Zhang, Y. Liu, Feature subset selection by gravitational search algorithm optimization, Information Sciences 281 (2014) 128–146.

[13] S. Das, A. Abraham, U. K. Chakraborty, A. Konar, Differential evolution using a neighborhood-based mutation operator, IEEE Transactions on Evolutionary Computation 13 (2009) 526–553.

[14] S. Dwivedi, M. Vardhan, S. Tripathi, Incorporating evolutionary computation for securing wireless network against cyberthreats, The Journal of Supercomputing (2020) 1–38.

[15] R. V. Rao, V. J. Savsani, D. Vakharia, Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Computer-Aided Design 43 (2011) 303–315.

[16] A. K. Shukla, P. Singh, M. Vardhan, An adaptive inertia weight teaching-learning-based optimization algorithm and its applications, Applied Mathematical Modelling 77 (2020) 309–326.

[17] E. Zorarpacı, S. A. Özel, A hybrid approach of differential evolution and artificial bee colony for feature selection, Expert Systems with Applications 62 (2016) 91–103.

[18] M. Allaoui, B. Ahiod, M. El Yafrani, A hybrid crow search algorithm for solving the dna fragment assembly problem, Expert Systems with Applications 102 (2018) 44–56.

[19] A. K. Shukla, S. K. Pippal, S. S. Chauhan, An empirical evaluation of teaching–learning-based optimization, genetic algorithm and particle swarm optimization, International Journal of Computers and Applications (2019) 1–15.

[20] S. Dwivedi, M. Vardhan, S. Tripathi, A. K. Shukla, Implementation of adaptive scheme in evolutionary technique for anomaly-based intrusion detection, Evolutionary Intelligence (2019) 1–15.

[21] E. Pashaei, E. Pashaei, N. Aydin, Gene selection using hybrid binary black hole algorithm and modified binary particle swarm optimization, Genomics (2018).

[22] A. K. Shukla, D. Tripathi, B. R. Reddy, D. Chandramohan, A study on metaheuristics approaches for gene selection in microarray data: algorithms, applications and open challenges, Evolutionary Intelligence (2019) 1–21.

[23] S. N. Qasem, A. Nazar, S. Attia Qamar, A learning based brain tumor detection system, CMC-COMPUTERS MATERIALS & CONTINUA 59 (2019) 713–727.

[24] A. K. Shukla, P. Singh, M. Vardhan, Neighbour teaching learning based optimization for global optimization problems, Journal of Intelligent & Fuzzy Systems 34 (2018) 1583–1594.

[25] S. Shahbeig, M. S. Helfroush, A. Rahideh, A fuzzy multi-objective hybrid tlbo-pso approach to select the associated genes with breast cancer, Signal Processing 131 (2017) 58–65.

[26] Z. Wang, R. Lu, D. Chen, F. Zou, An experience information teaching-learning-based optimization for global optimization., IEEE Trans. Systems, Man, and Cybernetics: Systems 46 (2016) 1202–1214.

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

[28] 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 (2012) 21–33.

[29] Y. Zhang, L. Wu, Y. Zhang, J. Wang, Immune gravitation inspired optimization algorithm, in: International Conference on Intelligent Computing, Springer, pp. 178–185.

[30] X. Zheng, W. Zhu, C. Tang, M. Wang, Gene selection for microarray data classification via adaptive hypergraph embedded dictionary learning, Gene 706 (2019) 188–200.

[31] S. Qummar, F. G. Khan, S. Shah, A. Khan, S. Shamshirband, Z. U. Rehman, I. A. Khan, W. Jadoon, A deep learning ensemble approach for diabetic retinopathy detection, IEEE Access 7 (2019) 150530–150539.

[32] Y. Xiao, J. Wu, Z. Lin, X. Zhao, A deep learning-based multi-model ensemble method for cancer prediction, Computer methods and programs in biomedicine 153 (2018) 1–9.

[33] Y. Wang, X. Li, R. Ruiz, Weighted general group lasso for gene selection in cancer classification, IEEE Transactions on Cybernetics (2018).

[34] J.-S. Wang, J.-D. Song, A hybrid algorithm based on gravitational search and particle swarm optimization algorithm to solve function optimization problems., Engineering Letters 25 (2017).

[35] L. E. A. d. S. Santana, A. M. de Paula Canuto, Filter-based optimization techniques for selection of feature subsets in ensemble systems, Expert Systems with Applications 41 (2014) 1622–1631.

[36] H.-C. Wu, X.-G. Wei, S.-C. Chan, Novel consensus gene selection criteria for distributed gpu partial least squares-based gene microarray analysis in diffused large b cell lymphoma (dlbcl) and related findings, IEEE/ACM transactions on computational biology and bioinformatics 15 (2017) 2039–2052.

[37] B. Xue, M. Zhang, W. N. Browne, X. Yao, A survey on evolutionary computation approaches to feature selection, IEEE Transactions on Evolutionary Computation 20 (2016) 606–626.

[38] A. Banitalebi, M. I. A. Aziz, Z. A. Aziz, A self-adaptive binary differential evolution algorithm for large scale binary optimization problems, Information Sciences 367 (2016) 487–511.

[39] M. Thakur, A new genetic algorithm for global optimization of multimodal continuous functions, Journal of Computational Science 5 (2014) 298–311.

[40] J. Kennedy, R. C. Eberhart, A discrete binary version of the particle swarm algorithm, in: Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on, volume 5, IEEE, pp. 4104–4108.

[41] S. Mirjalili, G.-G. Wang, L. d. S. Coelho, Binary optimization using hybrid particle swarm optimization and gravitational search algorithm, Neural Computing and Applications 25 (2014) 1423–1435.

[42] S. Mirjalili, S. Z. M. Hashim, A new hybrid psogsa algorithm for function optimization, in: Computer and information application (ICCIA), 2010 international conference on, IEEE, pp. 374–377.

[43] A. Nasir, M. O. Tokhi, Novel metaheuristic hybrid spiral-dynamic bacteria-chemotaxis algorithms for global optimisation, Applied Soft Computing 27 (2015) 357–375.

[44] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos, Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions, Neurocomputing 276 (2018) 2–22.

[45] A. Sharma, R. Rani, C-hmoshssa: Gene selection for cancer classification using multi-objective meta-heuristic and machine learning methods, Computer methods and programs in biomedicine 178 (2019) 219–235.

[46] I. Jain, V. K. Jain, R. Jain, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, Applied Soft Computing 62 (2018) 203–215.

[47] M. A. Al-Betar, M. A. Awadallah, Island bat algorithm for optimization, Expert Systems with Applications 107 (2018) 126–145.

[48] J. Feng, L. Jiao, F. Liu, T. Sun, X. Zhang, Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy, IEEE Transactions on Geoscience and Remote Sensing 53 (2015) 2956–2969.

[49] A. Özçift, A. Gülten, Genetic algorithm wrapped bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases, Digital Signal Processing 23 (2013) 230–237.

[50] A. K. Shukla, P. Singh, M. Vardhan, Hybrid tlbo-gsa strategy for constrained and unconstrained engineering optimization functions, Hybrid Metaheuristics: Research And Applications 84 (2018) 41.

[51] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE journal of biomedical and health informatics 21 (2016) 4–21.

[52] Y. Chen, Y. Li, R. Narayan, A. Subramanian, X. Xie, Gene expression inference with deep learning, Bioinformatics 32 (2016) 1832–1839.

[53] M. Mahmud, M. S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, IEEE transactions on neural networks and learning systems 29 (2018) 2063–2079.

[54] M. Tanveer, C. Gautam, P. Suganthan, Comprehensive evaluation of twin svm based classifiers on uci datasets, Applied Soft Computing 83 (2019) 105617.

[55] S. Mirjalili, A. Lewis, S-shaped versus v-shaped transfer functions for binary particle swarm optimization, Swarm and Evolutionary Computation 9 (2013) 1–14.

[56] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A.-Z. Ala'M, S. Mirjalili, Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems, Knowledge-Based Systems 145 (2018) 25–45.

[57] A. K. Shukla, Multi-population adaptive genetic algorithm for selection of microarray biomarkers, Neural Computing and Applications (2020) 1–22.

[58] P. E. Meyer, C. Schretter, G. Bontempi, Information-theoretic feature selection in microarray data using variable complementarity, IEEE Journal of Selected Topics in Signal Processing 2 (2008) 261–274.

[59] D. L. Tong, R. Mintram, Genetic algorithm-neural network (gann): a study of neural network activation functions and depth of genetic algorithm search applied to feature selection, International Journal of Machine Learning and Cybernetics 1 (2010) 75–87.

[60] H. Wang, X. Jing, B. Niu, A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data, Knowledge-Based Systems 126 (2017) 8–19.

[61] V. Bolón-Canedo, N. Sánchez-Maroño, A. Alonso-Betanzos, Distributed feature selection: An application to microarray data classification, Applied soft computing 30 (2015) 136–150.

[62] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Accelerating wrapper-based feature selection with k-nearest-neighbor, Knowledge-Based Systems 83 (2015) 81–91.

[63] M. Mollaee, M. H. Moattar, A novel feature extraction approach based on ensemble feature selection and modified discriminant independent component analysis for microarray data classification, Biocybernetics and Biomedical Engineering 36 (2016) 521–529.

[64] S. A. Medjahed, T. A. Saadi, A. Benyettou, M. Ouali, Kernel-based learning and feature selection analysis for cancer diagnosis, Applied Soft Computing 51 (2017) 39–48.

[65] L.-Y. Chuang, C.-H. Yang, C.-H. Yang, Tabu search and binary particle swarm optimization for feature selection using microarray data, Journal of computational biology 16 (2009) 1689–1703.

[66] J. Apolloni, G. Leguizamón, E. Alba, Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments, Applied Soft Computing 38 (2016) 922–932.

[67] J. Liu, X. Wang, Y. Cheng, L. Zhang, Tumor gene expression data classification via sample expansion-based deep learning, Oncotarget 8 (2017) 109646.

[68] R. Fakoor, F. Ladhak, A. Nazi, M. Huber, Using deep learning to enhance cancer diagnosis and classification, in: Proceedings of the international conference on machine learning, volume 28, ACM New York, USA.

[69] A. Sharma, R. Rani, An optimized framework for cancer classification using deep learning and genetic algorithm, Journal of medical imaging and health informatics 7 (2017) 1851–1856.

[70] W. J. Conover, R. L. Iman, Rank transformations as a bridge between parametric and nonparametric statistics, The American Statistician 35 (1981) 124–129.

List of Abbreviations	
TLBO	Teaching learning-based optimization
FS	Feature selection
GSA	Gravitational search algorithm
TLBOGSA	Teaching learning-based optimization gravitational search algorithm
BTLBOGSA	Binary teaching learning-based optimization gravitational search algorithm
mRMR	Minimum-redundancy-maximum-relevance
DNA	Deoxyribonucleic acid
PSO	Particle swarm optimization
GA	Genetic algorithm
ABC	Artificial bee colony
DE	Differential evolution
ACO	Ant colony optimization
BBHA	Binary black hole algorithm
SA	Simulated annealing
EA	Evolutionary algorithm
TS	Tabu search
WOA	Whale optimization algorithm
NTLBO	Neighbour teaching learning-based optimization
NB	Naïve Bayes
DT	Decision tree
SVM	Support vector machine
k-NN	k- nearest neighbor
CMIM	Conditional mutual information maximization
JMI	Joint mutual information
SRBCT	Small round blue-cell tumor
DLBCL	Diffuse large B-cell lymphoma

Alok Kumar Shukla: Conceptualization, Original draft preparation, Methodology, Software.

Pradeep Singh: Supervision, Writing- Reviewing and Editing.

Manu Vardhan: Visualization, Investigation, Validation.