IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 19, 2020, accepted June 25, 2020, date of publication July 1, 2020, date of current version July 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006197

Transformation Based Tri-Level Feature Selection Approach Using Wavelets and Swarm Computing for Prostate Cancer Classification

SUNIL KUMAR PRABHAKAR[1], (Member, IEEE), AND SEONG-WHAN LEE[2], (Fellow, IEEE)
[1]Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea
[2]Department of Artificial Intelligence, Korea University, Seoul 02841, South Korea

Corresponding author: Seong-Whan Lee (sw.lee@korea.ac.kr)

This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korean Government (MSIT), Department of Artificial Intelligence, Korea University, under Grant 2019-0-00079.

ABSTRACT Prostate Cancer is a cancer that occurs in the prostate- a small walnut shaped gland in men. This gland helps in the production of seminal fluid which is used to nourish and transport the sperm. One of the most common types of cancer in men is prostate cancer. A microarray dataset contains the microarray gene expression information. On a genome wide scale, gene expression profiles make it easy to analyze the patterns between genes and cancers, however the analysis of gene expression data is very difficult as it has a high dimensionality and low Signal to Noise Ratio (SNR). In this paper, a transformation-based Tri-level feature selection using wavelets for prostate cancer classification has been proposed. For the input microarray data, initially wavelets are applied and then the essential features are selected. Then the standardized gene selection techniques are implemented such as Relief-F, Fishers Score, Information Gain and SNR for a second level feature selection stage. Finally, before proceeding to classification, a third level feature selection by means of optimization techniques are implemented. The optimization techniques incorporated in this work are Marriage in Honey Bee Optimization Algorithm (MHBOA), Migrating Birds Optimization Algorithm (MBOA), Salp Swarm Optimization Algorithm (SSOA) and Whale Optimization Algorithm (WOA). This kind of an approach is totally new, and the best results show when SNR with WOA is classified with Artificial Neural Network (ANN) giving a classification accuracy of 99.48%. The second highest classification accuracy of 99.22% is obtained when Relief-F test with MBOA is classified with Naïve Bayesian Classifier (NBC).

INDEX TERMS Prostate cancer, feature selection, optimization, classification.

I. INTRODUCTION

The prostate is a very small walnut shaped gland present in the pelvis of men [1]. By getting a digital rectal exam, the prostate can be examined as it is located next to the bladder. A form of cancer that develops in the prostate gland is called prostate cancer [2]. Initially prostate cancer grows slowly and is confined to the prostate gland alone. Some kinds of prostate cancer can require a minimal treatment while others can spread very quickly and aggressively. While confined to the prostate gland and if it detected earlier, then the chances are better for a successful treatment [3]. The common risk factors of this disease include age, ethnicity, family history, smoking, diet etc. A few decades back, the cancer classification was based on clinical and morphological features and it had a very limited diagnostic ability [4]. Today with the advent of DNA microarray technology, a great progress in cancer classification has been stimulated as it is capable of monitoring thousands of gene expressions concurrently [5]. Therefore, to assess the expression patterns of thousands of genes, microarrays offer a very efficient technique of gathering the data. In a matrix format, the representation of the gene expression data is done where the genes are represented in the rows and the observation or the samples are represented by the columns [6]. In many classification problems, a big challenge is the high dimensionality of gene expression data [7]. The redundancy in the expressed data along with the very large number of features against a small sample size is one

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal.

of the main problem which leads to a huge computational complexity and also a less classification accuracy.

Prior to classification, feature selection is very important, and a comprehensive review of feature selection techniques discussed in literature is given in [8]. Only to avoid the curse of dimensionality problem, the feature selection techniques are needed and the mostly used strategies are filtering, wrapper, and embedded methods [9]. In the filtering methods, they do not incorporate learning as the subset features are selected in an independent manner from the learning classifiers. The main limitation of filtering techniques is that it assesses the individual features in isolation only and there are high chances to avoid the possible interaction among them. Sometimes, as a consequence of filtering technique, the procedure ends with highly correlated features and these features will definitely spoil the classification accuracy [10]. Later wrapper methods were developed that intended to wrap around a specific learning technique so that the final classifier is built based on the assessment of the selected feature subset in terms of the classification error estimation. The number of features can be reduced significantly, and the classification accuracy can be increased by wrapper methods however they too suffer from a high computational complexity and load [11]. Later to overcome the wrapper related problem, embedded techniques were developed so that the simultaneous feature selection along with a good learning classifier can be selected and it may suffer from computational load as well. However, the method approached in this work provides a better feature construction and ranking, enhances the search techniques, managing the versatility of the feature validity and a good clarity in the analysis of objective function aiding in the multivariate feature selection process.

Some important works discussed in literature about the prostate cancer classification is given below. The prostate cancer classification from the gene expression profiles by discriminant kernel-PLS was done by Tang *et al.* [12]. The molecular classification of prostate cancer by utilizing curated expression signatures was done by Markert *et al.* [13]. The diagnosis of prostate cancer using differentially expressed genes in stroma was performed by Jia *et al.* [14]. Machine learning techniques were implemented for classification and biomarker gene selection for prostate gene expression data by Ram *et al.* [15]. Spectral methods for prostate cancer classification using microarray data was analyzed by Kim [16]. From the clinical integration of inter-study microarray data, a robust prostate cancer marker gene was found by Xu *et al.* [17]. A case study on the recurrence of prostate cancer by utilizing machine learning techniques was done by Zupan *et al.* [18]. A combinatorial feature selection approach with ensemble neural network method for prostate classification of gene expression data was done by Liu *et al.* [19]. The delineation of prognostic biomarker in prostate cancer was done by Dhanasekaran *et al.* [20]. The selection of significant genes by using randomization test for prostate cancer classification using gene expression data was done by Mao *et al.* [21]. An optimization based prostate cancer classification from microarray gene expression data was done by Dangliyan *et al.* [22]. Prostate cancer diagnosis based on micro gene expression profiles with Linear Discriminant Analysis (LDA) was done by Bouazza *et al.* [23]. In this work, with the three-level feature selection utilization, wavelets with statistical tests and optimization techniques, the features were classified with five different classifiers. The organization of the work is as follows: The materials and methods are discussed in section 2 stating the application and importance of wavelet method and the standardized gene selection methods in our work. Section 3 engages about the four optimization techniques utilized to select the most important genes/features and section 4 explains the classification process followed by results and discussion in section 5 and conclusion in section 6.

II. MATERIALS AND METHODS

For the Prostate Cancer classification, a dataset was used which is publicly available online [24]. There are about 12600 genes here. There are 102 samples totally where Class 1 represents the tumor class with 52 samples and Class 2 represents the healthy class with 50 samples. The details of the dataset are tabulated in Table 1. The illustration of the work is found in Fig. 1.

TABLE 1. Dataset details.

Dataset	Number of genes	Class 1 (Tumor)	Class 2 (Healthy)	Total samples
Prostate Cancer	12600	52	50	102

A. APPLICATION OF WAVELETS

To explain a large set of data in an accurate manner, feature extraction is used as it is used to simplify the total amount of resources. For the analysis of high dimensional data, Discrete Wavelet Transform (DWT) is used widely [25]. In DWT, the approximation coefficients with some useful data from the high frequency coefficients jointly provides some useful information selected by the maximum modulus techniques. Assuming microarray data as a signal set, where the total number of genes is considered as the length of the signals and therefore signal processing techniques such as DWT are used here. Therefore, this work deals with the classification of specific microarray data into normal or abnormal based on DWT. In the fields of signal processing and image processing, DWT serves as a great multiresolution analysis tool. When using the wavelets here, the data is rearranged and gives a threshold to the wavelet coefficient. Then the approximate value of the raw data is calculated by means of implementation of an inverse function to the transformed data. The ranking of the wavelet coefficient is done, and the dominant feature is nothing but the top ranked wavelet coefficient. The top ranked wavelet coefficients are then sent to the other gene selection techniques for a secondary level feature selection process. The wavelet procedure is given in Algorithm 1.

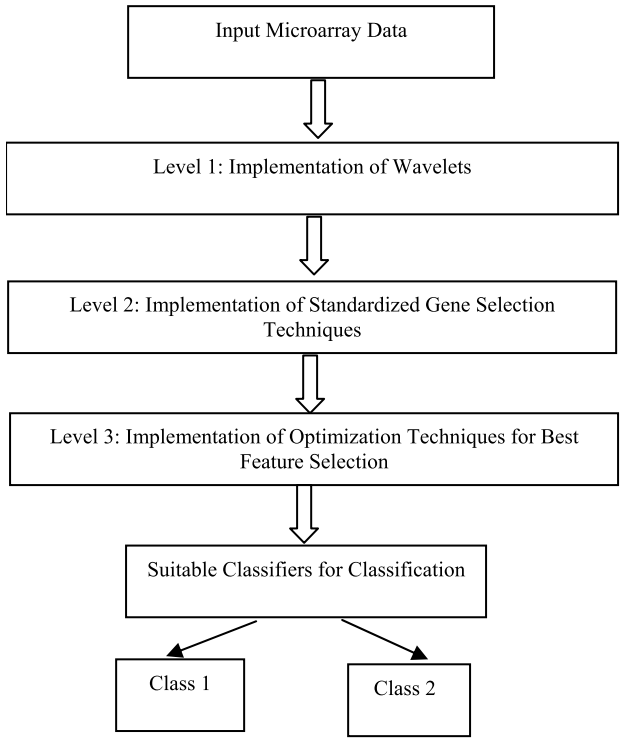

FIGURE 1. Illustration of the work.

As a result, from the 12,600 genes, the dominant 4000 genes were shortlisted after wavelet analysis and from this further important genes will be shortlisted through the standard gene selection techniques.

B. GENE SELECTION TECHNIQUES

The ranking criterion of this trilevel approach is considered here. The identification of a few top ranked features is done initially. The main intention of this particular aspect is to shortlist and extract the best 1000 Genes from 4000 Genes. Then the best genes from it will be selected once again from optimization techniques. The following gene ranking methods are explained here.

1) RELIEF F

A famous supervised feature weight approach or a feature rank approach, it primarily utilizes filter model and is used predominantly in most of the research works for both binary and multi class problems [9]. Here the worth of a feature is evaluated by means of repeatedly sampling a particular instance s. Then for the nearest instance of the same and different class, the value of the given feature is considered. For a binary class problem, the weight of a feature is considered as

$$R(g_x) = 2\sum_{t=1}^{s} d\left(g_{t,x} - g_{PQ(y_t),x}\right) - d\left(g_{t,x} - g_{PJ(y_t),x}\right) \quad (1)$$

where $g_{t,x}$, represents the value of y_t on gene g_x. $g_{PQ(y_t),x}$ represent the values on x^{th} gene to the closest point to y_t with a same class label. $g_{PJ(y_t),x}$ represent the values on x^{th} gene to the closest point to y_t with a different class label.

Algorithm 1 Wavelet Procedure

Input: dataset
Output: A subset comprising of dominant features of that dataset.
Method:

(1) Decomposition Step (S : $array\,[1, \ldots, n]$ of $reals$)
(2) for $j \leftarrow 1$ to $n/2$
(3) $S'[j] = (S[2j-1] + C[2j])/sqrt(2)$; // scaling coefficient
(4) $S'[n/2+j] \leftarrow (S[2j-1] - S[2j])/sqrt(2)$; // detail coefficient
(5) End for
(6) $S \rightarrow S'$
(7) Application of Threshold
(8) If $S[j] <= 0.5$ then $S[j] = 0$;//threshold value = 0.5
(9) Else $S[j] = S[j]$
(10) Inverse function implementation to approximate value is performed.
(11) Production of a new wavelet (with scaling and detail coefficients) is done here.
(12) Dominant features are selected and stored to be further reduced to a second level

2) FISHERS SCORE

A famous univariate supervised and feature weighting technique, this is used by many researchers [26]. Here, a rank to each feature is assigned in a manner that different values are assigned to samples for the different class and same values are assigned to samples from same class. The calculation of score or rank for each feature is done as follows:

$$Score(g_x) = \sum_{l=1}^{c} n_l\left(\mu_{x,l} - \mu_x\right)^2 / \sum_{l=1}^{c} n_l \sigma_{x,l}^2 \quad (2)$$

where for the gene g_x, the mean is represented as μ_x; $\mu_{x,l}$ and $\sigma_{x,l}$ indicate the mean and variance of gene g_x in class l. The number of class levels is expressed by l. In class l, the number of samples is represented by n_l.

3) INFORMATION GAIN

Based on a given attribute, due to the portioning of the examples there is an expected reduction in entropy, and it is called as Information Gain [27]. Here the information content of an attribute is represented by A which helps in representing a label. Therefore, to formulate the reduced data, depending on the selection criteria, very few top ranked features are obtained.

$$A = Info\,(Y) - Info_y \quad (3)$$

$$Info(Y) = -\sum_{x=1}^{l} P\,(c_x, Y) * \log\,(P\,(c_x, Y)) \quad (4)$$

$$Info_y(Y) = -\sum_{x=1}^{w} \frac{|W_x|}{|Y|} * Info\,(W_x) \quad (5)$$

where the number of classes is represented by l, for a gene y, the number of individual values is represented by w, the set of instances is represented as w_x whose values in gene y is equal to y_x.

4) SNR

Here the level of signal is evaluated with the level of noise and then the determination of the rank of each gene is done in this method [28]. A higher priority gene is represented by a higher SNR value, the selection of some of these genes are done as per this technique. The definition of SNR is as follows:

$$SNR_{(x)} = \frac{mean_{x1} - mean_{x2}}{std_{x1} + std_{x2}} \tag{6}$$

where the mean value of sample of class1 and class 2 is expressed as $mean_{x1}$ and $mean_{x2}$, the standard deviation of the samples in class 1 and class 2 is expressed by std_{x1} and std_{x2}. x represents the current genes under consideration.

III. OPTIMIZATION TECHNIQUES

The 1000 features selected through the standard genes selection techniques are further optimized with a third level feature selection by means of utilizing optimization techniques to select the top 50 – 150 genes. The application of the following optimization algorithms has been first of its kind to be utilized for prostate cancer classification.

A. MARRIAGE IN HONEY BEES OPTIMIZATION ALGORITHM

1) HONEY BEE COLONY

Among the insects, bees take the initial place that can be well modelled as swarm as it has the properties of swarm intelligence [29]. Three kinds of bees are present in a typical bee colony such as queen, male bee known as drone and female bee known as worker. The queen is the mother of the colony and is a couple of years old and she is only capable of laying eggs. From unfertilized eggs, drones are produced and are known as the father of the colony. Their number ranges are around a couples of hundreds. From the fertilized eggs, the worker bees are produced and every procedure such as feeding the queen and colony, management of broods, maintenance of combs, seeking and aggregation of food are done by these bees. Their number ranges from a few hundreds to many thousands. During the life of the queen bee, mating happens only once. The initiation of the mating flight begins with the dance of the queen. During the mating flight, the drones follow and mate with the queen. Depending on the speed and fitness of queen, the mating of a drone is done with the queen. The spermatheca of the queen is used to store the sperms of the drone and the gene pool of future generation is started here. Around two thousand fertilized eggs a day are laid by the queen and so it comes to around two hundred thousand a year. Once the discharge of the spermatheca is done, then she lays unfertilized eggs.

2) HONEY BEES OPTIMIZATION ALGORITHM

The explanation of the mating flight in the solution space can be done as the queen's acceptance of some drones she meets. At the start of the flight, the queen has a certain amount of energy and when her spermatheca is full or when her drops to a minimum level, then it turns back to the nest. The broods are generated after going back to the nest and with the help of worker bees crossover and mutation process, it is improved later. With the following probability of annealing functions, the mating of the drone with the queen bee happens and is represented as

$$prob\,fn\ (Queen, Dance) = e^{-difference/speed} \tag{7}$$

where (queen, drone) denotes the probability of the drone to be mingled to the spermatheca of the queen. $\Delta\,(fn)$ denotes the absolute difference between drone fitness and queen fitness. The speed of a queen at time t is represented as $S_p(t)$. This part represents the annealing function. The mating probability is high in cases where at first, the speed of queen is very high, or the drone fitness is equivalent to the queen fitness. Within the search space, the formulation of the time-dependent speed $S_p(t)$ and energy $E_n(t)$ of the queen in every pass is represented as:

$$\begin{aligned} S_p(t+1) &= \alpha \times S_p(t) \\ E_n(t+1) &= E_n(t) - \gamma \end{aligned} \tag{8}$$

Here, α is a factor and it belongs to [0,1]. In every pass, the amount of energy reduction is represented as γ. The original MHBO algorithm is shown as follows in Algorithm 2.

B. MIGRATING BIRDS OPTIMIZATION ALGORITHM

It is a very famous nature-inspired metaheuristic algorithm based on V-formation flight of the migration birds [30]. A set of individuals is present in this algorithm where every set is related to a solution and named as birds in MBO. In a V shaped formation, the individuals are aligned, where the first individual is corresponding to the leader bird in the flock. By generating a number of neighbour solutions, the leader attempts to improve itself in MBO. Then the number of its own neighbours are evaluated by the individual during V formation. From the previous individual, a total number of best discarded neighbour solutions are also evaluated. When a solution leads to an improvement, then it is replaced to the older solution so that maximum improvement is yielded. The process is repeated for $iter_{\max}$ once all the individuals have been analyzed. The lead individual is progressed to the end of one of the lines of the V formation once these sufficient iterations are reached. Then one of its direct follower individual becomes the new leader of the flock. For another $iter_{\max}$ iterations, the process is restored so that a new formation is obtained. With a given stopping criteria is fulfilled, till then the complete MBO is carried out. The generation order is an important factor for assessing the initial positions of the individuals along the V formation. The generation of the first individual will be the leader bird of the flock. The direct

Algorithm 2 MHBOA

Initialization of workers
Random generation of queens
Implement local search to obtain a good queen
For a specific and predefined number of mating-flights
For every queen in the queen list
Initialize parameters such as Energy (E_n), Position and Speed (S_p)
Movement of queen between states
Choose drone probabilistically $prob\,fn\ (Queen, Drone) = e^{-(d/s)}$
If selection of drone is done, then
Addition of sperms to the spermatheca of queen

$$S_p(t+1) = \alpha \times S_p(t)$$
$$E_n(t+1) = E(t) - \gamma$$

End if
Updation of the speed and internal energy of queen
End for each
Produce broods by crossover and mutation
Utilize works to brood improvement
Updation of fitness of workers
While the best brood is better than the worst queen
Replacement of the least fit queen by best brood
Elimination of the worst brood
End While
End for

followers will be the second and third individuals generated. Once when the initial population is generated, the organization of the individuals into a V shaped formation is done. For the MBO algorithm, the input parameters defined by the user are as follows:

t_{birds} : number of individuals named as birds

K : maximum number of neighbour solutions merged by individuals

$iter_{\max}$: before changing the leader individuals, the number of iteration count

λ : random number count generated by each individual

δ : number of discarded solutions to be shared among the individuals.

Initially, the t_{birds} are generated individually. The number of generated solutions g_s by the population P is set to 0. The MBO search process starts once the generation of population is done. The generation of λ random neighbours solutions are done by the leader individual during the search process by means of utilizing a neighbourhood structure. When the best solution is generated and when it leads to a higher improvement in the objective function value, then the neighbour solution is used to replace the solution related to the neighbour. By means of utilizing a neighbourhood structure, the generation of $\lambda - \delta$ neighbour solutions are done by each direct follower individuals. The best δ discarded solutions are also received from every individual in front of it. If the generated or received solutions leads to an improvement of the solution, then it replaces the individual set. Unless a prefixed number of iterations are reached, the V-formation is maintained. Once it is achieved, the last solution is replaced by the leader individual and the new leader is chosen to form one of its direct follower individually. For other $iter_{\max}$ iterations, the search process is restarted. Unless a number of neighbour solutions M have been generated with the help of search process, the execution of the MBO search process is done. The procedure is explained in Algorithm 3 as follows:

Algorithm 3 MBOA

Generate t_{birds} initial solution randomly and it is arranged on a V formation arbitrarily.
$g_s = 0$
While $g_s < M$ do
for $(q = 0, q < iter_{\max}; q++)$ *do*
Leading solution improvement by evaluating λ neighbours
$g_s = g_s + \lambda$
For all (solutions s in the flock except lead) do
Leading solution improved by generating and evaluation $\lambda - \delta$
$g_s = g_s + (\lambda - \delta)$
end for
end for
Movement of the leading solution to end
Follow one of the solutions to the leader position
End while
Return best solution in the flock

C. SALP SWARM OPTIMIZATION ALGORITHM

Salps belong to the salpidae family and their body is designed in a limpid cylindrical manner [31]. When their movement and texture is considered, they look similar to jellyfishes. The bodies of salps move forward due to the movement push of water. Maintaining salps in laboratory environment is externally hard and because of its hardly accessible living environment, the biological research about them is still in early stages. To build the salp swarm algorithm, the main inspiration is the salps swarming attitude. A salp chain is nothing but a composition of salp swarm found in profound oceans. During the foraging methodology, a better locomotion is achieved using this chain. For the formulation of mathematical model for the scalp chains, the entire salp population is divided into two groups such as heads and followers. At the beginning of the chain is present head position while the rest of the chain is considered as followers. Like other swarm-based methods, the determination of salp

location is done by means if considering of an n-dimensional search area. The n represents a presented problem where a number of variables are considered inside. To reserve the position of all salps, a two-dimensional matrix indicated as y is considered. The food source F is assumed as the target of the swarm in the search space. To upgrade the location of leader, the suggestion of the following equation is done as;

$$y_k^1 = \begin{cases} F_k + r_1 \left((u_k - l_k) r_2 + l_k \right), & r_3 \geq 0 \\ F_k - r_1 \left((u_k - l_k) r_2 + l_k \right), & r_3 < 0 \end{cases} \tag{9}$$

where in the k^{th} dimension y_k^1 gives the position of the first salp (leader).

In the k^{th} dimension, the food source position is represented as F_k, the upper bound of the k^{th} dimension is specified as U_k and the lower bound of the k^{th} dimension is expressed as l_k, r_1, r_2 and r_3 are the random numbers. The position is only updated by the leader with respect to the food source as shown in (9). The behaviour of exploration and exploitation is done by the coefficient r_1 as it is most significant in SSOA and represented as

$$r_1 = 2e^{-(4m/M)^2} \tag{10}$$

where the current iteration is expressed as m and the maximum number of iterations is expressed by M. In the interval of [0,1] the parameters r_2 and r_3 are uniformly generated in a random manner. They take complete control in determining the upcoming position in k^{th} dimension and decides whether it should progress towards the positive infinity or negative infinity. The following equation is utilized in order to update the position of followers as:

$$y_k^j = \frac{1}{2} s t^2 + w_0 t \tag{11}$$

where $j \geq 2$, y_k^j gives the position of j^{th} followers scalp in k^{th} dimension, the time is expressed as $'t'$, w_0 is the initial speed, $s = w_{final}/w_0$ where $w = y - y_0/t$

The time process considered in this optimization is an iteration and the discrepancy between iteration is equal to 1, and when assuming $w_0 = 0$, the equation is expressed as follows

$$y_k^j = \frac{1}{2} \left(y_k^j + y_k^{j-1} \right) \tag{12}$$

where $j \geq 2$ and the position of j^{th} follow salp in k^{th} dimension is expressed by y_k^j. Further details can be found in Pseudocode 1.

Pseudocode 1: Pseudocode of the SSOA algorithm

```
Initialize the salp population y_j (j = 1, 2, ..., n) considering
u_b and l_b
While (end condition is not satisfied)
Compute the fitness for every search agent
F is the best search agent
Update r_1 by means of using velocity and position updation
        For each salp (y_j)
                If (j == 1)
                        Update the position of the leading
salp using velocity updation equation
                Else
                        Update the position of the fellow
salp by global search equation
                End
        End
        Modify the salps based on bounds (both upper and
lower)
End
Return F
```

D. WHALE OPTIMIZATION ALGORITHM

One of the biggest mammals on planet is whales. Vital classification of this animal includes killer, blue, humpback and finback [32]. From the seas and oceans, they need to breathe and therefore they do not sleep. Also, in a whale, only half of the brain can work. The living of the whale can be done either alone or in social groups, for instance killer whales usually thrive as a family. The food habits of whale include fishes and krill species mainly for most of their diet. There are significant cells in the most important regions of their brain. For controlling the feelings, emotion, for judgments and interaction, these cells are of much use. Similar to humans, whales have a lot of brain cells accounting for their smartness levels. The whales can think, learn, communicate, interact emote just like a human does. Even a kind of dialect can be developed by the whales. The hunting aspect of humpback whale is considered special as they use a technique called bubble-net feeding method. The following 3 phases are included in the WOA.

(i) Encircling the prey phase
(ii) Bubble net attacking phase
(iii) Searching for prey phase

(i) ENCIRCLING THE PREY PHASE

The objective prey is considered as the optimal candidate solution in the WOA. The encircling of the prey by the whales is expressed as follows:

$$P = \left| M \vec{Y}^*(t) - Y(t) \right| \tag{13}$$

$$Y(t+1) = \vec{Y}^*(t) - \vec{B}.\vec{P} \tag{14}$$

where the current position iterations is referred by t, the coefficient vectors are $\vec{B}$ and $\vec{M}$. The position vectors of the present optimal solution is done by $\vec{Y}^*$. The position vector is expressed as $\vec{Y}$, $||$ indicates the absolute value and $(\bullet)$ represents the element by element multiplication. Using the following mathematical form, the vectors $\vec{B}$ and $\vec{M}$ are evaluated as

$$\begin{aligned} \vec{B} &= 2.\vec{b}.\vec{r} - \vec{b} \\ \vec{M} &= 2.\vec{r} \end{aligned} \tag{15}$$

where $\vec{r}$ is nothing but the random number in the interval of [0, 1]. At this stage, it implies that the whale is going to attack the prey by means of using bubble net.

(ii) BUBBLE NET ATTACKING PHASE

To assess the mathematical model of the bubble net phase of the whales, two techniques are proposed as follows:

(i) Shrinking encircling technique: In this technique, the value of $\vec{b}$ is reduced and $\vec{B}$ is assumed as a randomly selected value in the interval [-b, b] such that there is a reduction in the value of b from 2 to 0. $\vec{B}$ is considered to have random values in the interval [−1,1].

(ii) Spiral updating position technique: For assessing the position of both whale and prey, this spiral equation can be used as

$$\vec{Y}(t+1) = \vec{P}'.e^{c}\cos(2\pi n) + \vec{Y}^{*}(t) \tag{16}$$

where $\vec{P}' = \left|\vec{Y}^{*}(t) - \vec{Y}(t)\right|$ explains the distance of the j^{th} whale to the prey.

The constant is denoted by c and the random number by n expressed in the interval of [−1, 1]. (•) is indicated by the element-by-element multiplication. Around the prey, the whales start surrounding in a shrinking circle with 0.75% probability to choose between spiral model or shrinking model to that the whale position can be updated. For this behaviour, the mathematical model is as follows:

$$\vec{Y}(t+1) = \begin{cases} \vec{Y}^{*}(t) - \vec{B}.\vec{P} & if\ r < 0.5 \\ \vec{P}'.e^{c}.\cos(2\pi n) + \vec{Y}^{*}(t) & if\ r \geq 0.5 \end{cases} \tag{17}$$

where r is expressed as a random value in the range of [0,1].

(iii) PREY PHASE SEARCHING

Here, $\vec{B}$ is defined as the random value between 1 and −1. To perform the global search and to enable this algorithm $\left|\vec{B}\right| > 1$ is considered. By using the following mathematical equation, this equation can be represented as

$$\vec{P} = \left|\vec{M} \cdot Y_{rand} - \vec{Y}\right|$$
$$\vec{Y}(t+1)\vec{Y}_{rand} - \vec{B}.\vec{P} \tag{18}$$

where the random position vectors is represented as $\vec{Y}_{rand}$. Utilizing the solutions that have been randomly selected, the searching process is initiated by this algorithm. By choosing the search agents, their appropriate positions are chosen randomly with each iteration. The selection of the random search agent is done when $\left|\vec{B}\right| > 1$ and the range of parameters should be 2 to 0.

IV. CLASSIFICATION PROCEDURES

The following classifiers are utilized here for classification of the optimized genes or features. The classifiers used here are Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Naïve Bayesian Classifier (NBC) and Support Vector Machine (SVM).

A. ARTIFICIAL NEURAL NETWORK

One of the important tools utilized in Machine Learning is ANN [33]. To replicate the ways similar to human learning, this brain inspired system was developed. ANN consists of input, hidden and output layers, where the use of hidden layer is to transform the input into something productive and useful that the output layer can use. Here the number of hidden layers is varied from 1 to 100 to get the better classification accuracy with minimum Mean Square Error (MSE).

B. LDA

One of the most famous supervised learning technique is LDA and is used in Machine Learning paradigm to seek a linear combination of features that can classify into 2 or more classes [34]. In LDA every class or collection has an equal covariance matrix. The classification in LDA is done in such a way that the separation between the classes is maximized and the separation with the classes is minimized. The formulation of the LDA is expressed as:

$$Z^{T}\Sigma^{-1}\left(\bar{Z}_2 - \bar{Z}_1\right) - \frac{1}{2}\left(\bar{Z}_2 + \bar{Z}_1\right)^{T}\Sigma^{-1}\left(\bar{Z}_2 - \bar{Z}_1\right) > t \tag{19}$$

where the data matrix is represented as Z, $\bar{Z}_1$ and $\bar{Z}_2$ are the mean vectors for the two groups, Σ represents the sample covariance matrix, Z^{T} represents the transpose of the data matrix Z, t represents the threshold of the decision boundary. The value of t may be greater than, less than or equivalent to zero. The classes are similar if the values of t is equal to 0 and it is classified into the cancer group of the value of $t > 0$.

C. NBC

They belong to group of simple probabilistic classifiers. Based on Baye's theorem, the collection of classification algorithms was performed [35]. A common principle is shared here such that every feature has equal contribution to the net outcome and for the classification the pair of features being used is independent of each other. Generally the feature matrix (z_i) containing the explanatory variables and the response vector (x) having the outcomes are the two important parts in this classifier. Bayes theorem is as follows:

$$P(x|z_1, \ldots, z_n) = \frac{P(x)\prod_{i=1}^{n} P(Z_i|x)}{\prod_{i=1}^{n} P(Z_i)} \tag{20}$$

where the conditional probability of x given $z_1, z_2, \ldots, z_n$ is expressed as $P(x|z_1, \ldots, z_n)$

The probability of x is $P(x)$, and the conditional probability of z_i $(i = 1, 2, 3, \ldots, n)$ given x is represented as $P(z_i|x)$, $P(z)$ represents the probability of z_i, Product symbol is represented as π. The output with maximum probability is selected by analyzing the probability of the specific set of inputs for all the possible values of class variable x. Mathematically, it is represented as;

$$x = \arg\max_{x} P(x)\prod_{i=1}^{n} P(z_i|n) \tag{21}$$

4.4 SVM: It is also a famous supervised learning technique that is used in Machine Learning [36]. A dataset D consists of

TABLE 2. Performance analysis of classifiers in terms of classification accuracies with MHBOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	95.698969	82.29	94.2725	86.71875
	100	98.96	95.315	95.835	94.795
	150	91.1475	88.55	95.315	88.55
LDA	50	86.71875	94.2725	84.375	81.25
	100	93.75	96.875	93.23	97.26656
	150	96.875	89.6	91.1475	98.4375
NBC	50	82.29	83.594	85.42	96.875
	100	81.25	92.71	92.71	94.795
	150	92.71	96.355	91.67	93.23
SVM (with RBF Kernel)	50	98.4375	97.915	91.1475	91.67
	100	90.1125	96.355	93.23	87.10938
	150	97.386875	79.17	90.1125	97.3975
Average		**92.111424**	**91.08345833**	**91.53875**	**92.34122**

a series of variables $v_1, v_2, \ldots, v_n \in V$ is highly associated with the observation. The learning of a function $f : Z \rightarrow v$ from D is done by this separating hyper plane classifier method so that prediction of the label of a new observation $z \in Z$ by $f(Z)$ is done and therefore it is finally classified into 2 respective classes as $v \in \{-1, +1\}$. A separating hyperplane has the property that

$$W^T Z + b > 0 \quad if\ v_i = 1 \tag{22}$$

$$W^T Z + b < 0 \quad if\ v_i = -1 \tag{23}$$

A factor called margin is used here to assess the minimal distance from the observation to the hyperplane. The margin is usually large for the maximal margin hyperplane and is called as separating hyperplane also. SVM utilizes kernel trick if the data cannot be separated in a linear manner. For the projection of low dimensional input space to a high dimensional space, kernels are utilized. In other words, a non-separable problem is converted to a separable problem.

V. RESULTS AND DISCUSSION

It is classified with a 10-fold cross validation method and the performance of it is shown in tables below. The mathematical formulae for computing the Performance Index (PI), Sensitivity, Specificity and Accuracy is mentioned in literature and using the same, the values are computed and exhibited [37]. PC is Perfect Classification; MC is Missed Classification and FA is False Alarm in the expressions below.

The sensitivity is computed as

$$Sensitivity = \frac{PC}{PC + FA} \times 100 \tag{24}$$

Specificity is computed as

$$Specificity = \frac{PC}{PC + MC} \times 100 \tag{25}$$

Accuracy is expressed as

$$Accuracy = \frac{Sensitivity + Specificity}{2} \tag{26}$$

Performance Index (PI) is expressed as

$$PI = \left(\frac{PC - MC - FA}{PC}\right) \times 100 \tag{27}$$

Table 2 shows the Performance Analysis of Classifiers in terms of Classification Accuracies with MHBOA for different gene selection techniques using 50-100-150 selected genes. As indicated in the Table 2 that ANN with 100 genes selected for Relief –F gene selection Method achieved high accuracy of 98.96%. Low accuracy value of 79.17% is reached at Fishers scores gene selection in SVM Classifier with RBF Kernel for 150 selected genes. SNR gene selection technique attained at high average accuracy of 92.34% across the classifiers.

Table 3 indicates the Performance Analysis of Classifiers in terms of Classification Accuracies with MBOA for different gene selection techniques using 50-100-150 selected genes. As reported in the Table 3 that NBC classifier with 100 genes selected for Relief –F gene selection Method achieved high accuracy of 99.22%. Whereas Low accuracy value of 75.5% is reported for information gain gene selection in ANN Classifier 100 selected genes. Fishers score gene selection technique attained at high average accuracy of 91.36% across the classifiers.

TABLE 3. Performance analysis of classifiers in terms of classification accuracies with MBOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	92.71	83.594	92.19	93.75
	100	90.625	96.355	75.5	85.67875
	150	81.77	97.395	78.6475	77.73313
LDA	50	94.2725	92.71	81.38	93.23
	100	80.21	95.835	84.375	90.625
	150	85.42	91.67	89.6	89.6
NBC	50	85.9375	94.795	94.01125	85.67875
	100	99.22	80.73	96.355	75.53125
	150	94.321484	83.594	86.328125	95.835
SVM (with RBF Kernel)	50	93.23	94.795	95.835	91.1475
	100	95.315	91.1475	85.67875	95.315
	150	93.75	93.75	77.733125	91.1475
Average		**90.565124**	**91.36420833**	**86.46947917**	**88.77266**

TABLE 4. Performance analysis of classifiers in terms of classification accuracies with SSOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	97.395	82.29	91.67	95.835
	100	84.63625	98.17625	92.71	84.63625
	150	78.451563	98.96	92.71	80.07938
LDA	50	75.75	97.915	79.95	95.055
	100	94.01125	92.45	75.75	94.53375
	150	91.1475	89.3375	91.1475	98.96
NBC	50	80.73	85.9375	86.71875	92.19
	100	95.315	92.19	93.49	91.93
	150	75.53125	94.141875	95.315	85.9375
SVM (with RBF Kernel)	50	78.255625	93.75	94.53375	**75.5**
	100	84.375	90.625	98.96	84.375
	150	97.915	93.62	90.1125	82.29
Average		**86.12612**	**92.44942708**	**90.255625**	**88.44349**

Table 4 denotes the Performance Analysis of Classifiers in terms of Classification Accuracies with SSOA for different gene selection techniques using 50-100-150 selected genes. As observed in the Table 4 that ANN classifier with 150 genes selected for Fishers score gene selection Method, SVM Classifier with 100 genes selected by Information gain techniques and LDA Classifier with 150 gene selected at SNR gene selection method achieved high accuracy of 98.96%. Low accuracy value of 75.5% is achieved for SNR gene selection in SVM Classifier 50 selected genes. Fishers score gene selection technique peaked at high average accuracy of 92.449% across the classifiers.

TABLE 5. Performance analysis of classifiers in terms of classification accuracies with WOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	82.616	91.67	97.655	92.45
	100	78.6475	88.878125	84.17975	99.48
	150	87.5	90.1125	95.315	94.53375
LDA	50	90.625	93.75	76.1121875	95.055
	100	92.71	98.96	95.835	91.67
	150	80.73	85.42	97.915	90.625
NBC	50	78.38625	90.88625	80.99	88.18906
	100	91.865	92.71	95.055	92.19
	150	97.655	77.923125	92.45	98.96
SVM (with RBF Kernel)	50	89.3375	92.71	94.2725	92.71
	100	86.71875	96.875	84.63625	95.055
	150	97.655	98.96	80.37203125	89.3375
Average		**87.8705**	**91.57125**	**89.56564323**	**93.35461**

TABLE 6. Performance analysis of classifiers in terms of classification performance index with MHBOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	89.75052	45.16	87.04	63.765
	100	97.87	89.98	91.58	88.38
	150	78.465	72.795	89.98	72.795
LDA	50	63.765	87.04	54.54	40
	100	85.7	93.33	84.315	94.07187
	150	93.33	78.93	78.465	96.76
NBC	50	45.16	51.163	58.83	93.33
	100	40	82.93	82.93	88.38
	150	82.93	92.455	80.01	84.315
SVM (with RBF Kernel)	50	96.76	95.65	78.465	80.01
	100	77.925	92.455	84.315	65.2125
	150	94.47187	28.59	77.925	94.725
Average		**78.84395**	**75.87316667**	**79.0329167**	**80.14536**

Table 5 reveals the Performance Analysis of Classifiers in terms of Classification Accuracies with WOA for different gene selection techniques using 50-100-150 selected genes. As mentioned in the Table 5 that ANN classifier with 100 genes selected for SNR gene selection Method achieved high accuracy of 99.48%. Low accuracy value of 76.112% is ebbed for information gain gene selection in LDA Classifier with 50 selected genes. SNR gene selection technique attained at high average accuracy of 93.35% across the classifiers.

Table 6 indicates the Performance Analysis of Classifiers in terms of Performance Index with MHBOA for different gene selection techniques using 50-100-150 selected genes. As shown in the Table 6 that ANN Classifier with

TABLE 7. Performance analysis of classifiers in terms of classification performance index with MBOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	82.93	51.163	81.47	85.7
	100	76.92	92.455	3.92	59.85
	150	42.58	94.49	25.405	19.58625
LDA	50	87.04	82.93	40.645	84.315
	100	34.295	91.58	54.54	76.92
	150	58.83	80.01	78.93	78.93
NBC	50	60.87	88.38	86.37	59.85
	100	98.4025	36.875	92.455	4.155625
	150	87.18578	51.163	62.3175	91.58
SVM (with RBF Kernel)	50	84.315	88.38	91.58	78.465
	100	89.98	78.465	59.85	89.98
	150	85.7	85.7	19.58625	78.465
Average		**74.08736**	**76.79925**	**58.0890625**	**67.31641**

TABLE 8. Performance analysis of classifiers in terms of classification performance index with SSOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	94.49	45.16	80.01	91.58
	100	55.6125	96.205	82.93	55.6125
	150	24.14938	97.87	82.93	33.49875
LDA	50	5.805	95.65	32.7325	89.18
	100	86.37	82.2	5.805	87.71
	150	78.465	77.39625	78.465	97.87
NBC	50	36.875	60.87	63.765	81.47
	100	89.98	81.47	85.0075	80.74
	150	4.155625	86.705	89.98	60.87
SVM (with RBF Kernel)	50	23.01625	85.7	87.71	3.92
	100	54.54	76.92	97.87	54.54
	150	95.65	85.35375	77.925	45.16
Average		**54.0924**	**80.95833333**	**72.0941667**	**65.17927**

100 genes selected for Relief –F gene selection Method achieved high Performance Index of 97.87%. Low Performance Index of 28.59% is arrived at Fisher s scores gene selection in SVM Classifier with RBF Kernel for 150 selected genes. SNR gene selection technique attained at high average Performance Index of 80.14% across the classifiers.

Table 7 shows the Performance Analysis of Classifiers in terms of Performance Index with MBOA for different gene selection techniques using 50-100-150 selected genes. As identified in the Table 7 that NBC classifier with 100 genes selected for Relief –F gene selection Method achieved high Performance Index of 98.402%. Whereas Low Performance

TABLE 9. Performance analysis of classifiers in terms of classification performance index with WOA for different gene selection techniques using 50 – 100-150 selected genes.

Classifiers	Number of Genes selected	Gene Selection Techniques			
		Relief - F	Fishers Score	Information Gain	SNR
ANN	50	46.66075	80.01	95.07	82.2
	100	25.405	74.7121875	53.69575	98.935
	150	66.66	77.925	89.98	87.71
LDA	50	76.92	85.7	8.255	89.18
	100	82.93	97.87	91.58	80.01
	150	37.1475	58.83	95.65	76.92
NBC	50	23.8125	77.6925	38.57375	70.68609
	100	80.5575	82.93	89.18	81.47
	150	95.07	20.505	82.2	97.87
SVM (with RBF Kernel)	50	77.39625	82.93	87.04	82.93
	100	63.765	93.33	55.6125	89.18
	150	95.07	97.87	35.1240625	77.39625
Average		**64.28288**	**77.52539063**	**68.4967552**	**84.54061**

TABLE 10. Consolidated average performance analysis of classifiers in terms of classification accuracy performance index with four optimization techniques for average of gene selection techniques using 50-100-150 selected genes for Prostate cancer.

Classifiers	Optimization Method	MHBOA		MBOA		SSOA		WOA	
	Parameters / Number of Genes selected	Accuracy (%)	Performance Index (%)	Accuracy (%)	Performance Index (%)	Accuracy (%)	Performance Index (%)	Accuracy (%)	Performance Index(%)
ANN	50	89.74505	71.42888	90.561	75.31575	91.7975	77.81	91.09775	75.98519
	100	96.22625	91.9525	87.03969	58.28625	90.03969	72.59	87.79634	63.18698
	150	90.89063	78.50875	83.88641	45.51531	87.55023	59.61203	91.86531	80.56875
LDA	50	86.65406	61.33625	90.39813	73.7325	87.1675	55.84188	88.88555	65.01375
	100	95.28039	89.35422	87.76125	64.33375	89.18625	65.52125	94.79375	88.0975
	150	94.015	86.87125	89.0725	74.175	92.64813	83.04906	88.6725	67.13688
NBC	50	87.04475	62.12075	90.10563	73.8675	86.39406	60.745	84.61289	52.69121
	100	90.36625	73.56	87.95906	57.97203	93.23125	84.29938	92.955	83.53438
	150	93.49125	84.9275	90.01965	73.06157	87.73141	60.42766	91.74703	73.91125
SVM (with RBF Kernel)	50	94.7925	87.72125	93.75188	85.685	85.50984	50.08656	92.2575	82.57406
	100	91.70172	79.97688	91.86406	79.56875	89.58375	70.9675	90.82125	75.47188
	150	91.01672	73.92797	89.09516	67.36281	90.98438	76.02219	91.58113	76.36508
Average		**91.76871**	**78.47385**	**89.29287**	**69.07302**	**89.31867**	**68.08104**	**90.5905**	**73.71141**

Index value of 3.92% is demonstrated for information gain gene selection in ANN Classifier 100 selected genes. Fishers score gene selection technique attained at high average Performance Index of 76.79% across the classifiers.

Table 8 depicts the Performance Analysis of Classifiers in terms of Performance Index with SSOA for different gene selection techniques using 50-100-150 selected genes. As indicated in the Table 8 that ANN classifier with 150 genes selected for Fishers score gene selection Method gives the lowest PI. SVM Classifier with 100 gene selected by Information gain techniques and LDA Classifier with 150 gene selected at SNR gene selection method achieved high Performance Index of 97.87%. Low Performance Index value of 3.92% is ebbed for SNR gene selection in SVM Classifier with 50 selected genes. Fishers score gene selection technique peaked at high average Performance Index of 80.95% across the classifiers.

Table 9 displays the Performance Analysis of Classifiers in terms of performance Index with WOA for different gene selection techniques using 50-100-150 selected genes. As Shown in the Table 9 that ANN classifier with 100 genes selected for SNR gene selection Method achieved high Performance Index of 98.935%. Low Performance Index of 8.255% is achieved for information gain gene selection in LDA Classifier with 50 selected genes. SNR gene selection technique attained at high average Performance Index of 84.54% across the classifiers.

Table 10 shows the Consolidated Average Performance Analysis of Classifiers in terms of Classification Accuracy, Performance Index with four optimization techniques

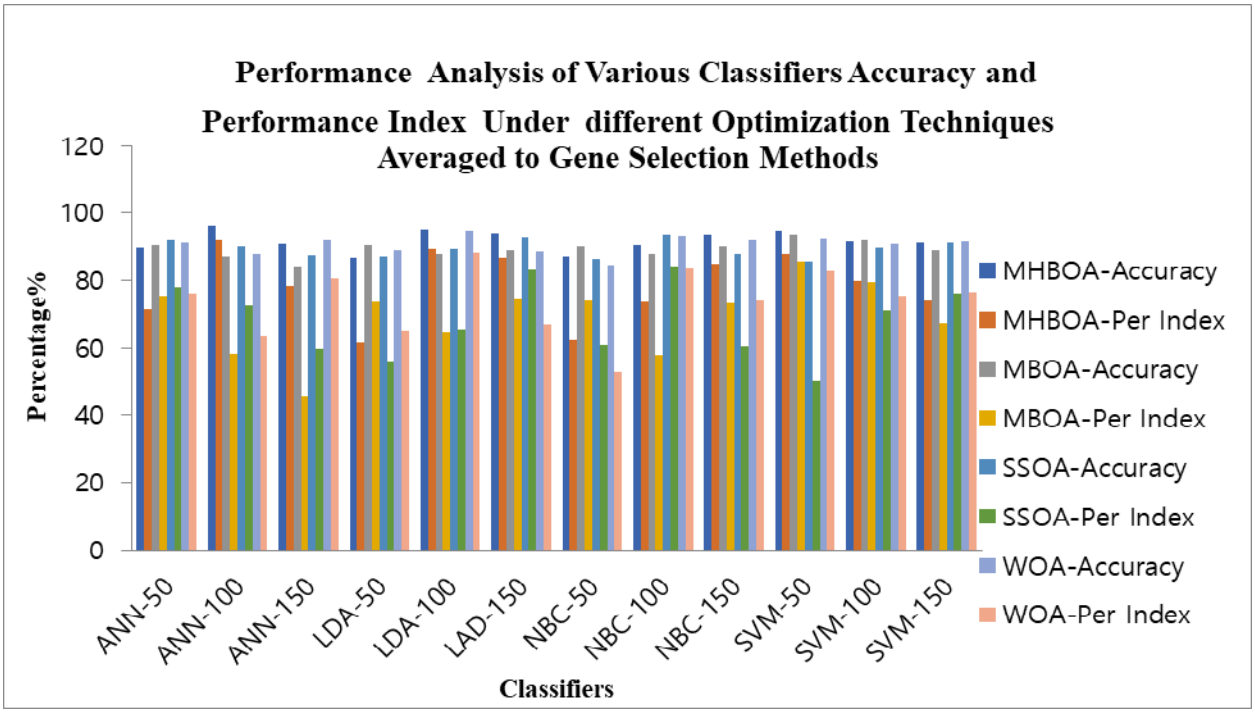

FIGURE 2. Performance analyses of various classifiers accuracy and performance index under different optimization techniques averaged to gene selection methods.

for Average of gene selection techniques using 50-100-150 selected genes for Prostate Cancer. As depicted in the Table 10 that the highest average accuracy 96.22% is attained at ANN classifier with 100 genes selected for MHBOA optimization method. 91.95% of Performance Index is also reported for the same classifier. Once again ANN classifier with 150 selected genes for MBOA optimization method reaches low accuracy of 83.88% and its corresponding Performance Index of 45.51%. MHBOA optimization techniques outperformed all other optimization methods in terms of high average accuracy and Performance Index across the classifiers.

Fig. 2. displays the Consolidated Average Performance Analysis of Classifiers in terms of Classification Accuracy, Performance Index with four optimization techniques for Average of gene selection techniques using 50 – 100-150 selected genes for Prostate Cancer. As demonstrated in the Fig. 2. that the highest average accuracy 96.22% is attained at ANN classifier with 100 genes selected for MHBOA optimization method. 91.95% of Performance Index is also reported for the same classifier. Once again ANN classifier with 150 selected genes for MBOA optimization method reaches low accuracy of 83.88% and its corresponding Performance Index of 45.51%. MHBOA optimization techniques outperformed all other optimization methods in terms of high average accuracy and Performance Index across the classifiers.

The best results are projected when SNR with WOA is classified with ANN giving a classification accuracy of 99.48%. The main reason for such a performance would be due to the intrinsic property of the WOA algorithm where it has the vital capability of exploring the mechanism of position updation easily, thereby the exploitation ability is enhanced and the local minima could be eliminated quickly. When such a combination is utilized to classify with ANN, a classification accuracy of 99.48% is obtained. A second highest classification accuracy of 99.22% is obtained when MBOA algorithm is classified by NBC and it is due to the production of broods by crossover and mutation state in MBOA algorithm, and so the fitness of every worker can be improved easily in the best possible manner and the worst brood is eliminated thereby giving the best possible genes and therefore when classified with NBC which specializes in Bayesian characteristics, this classification accuracy is achieved. The computation complexity for all the classifiers among four gene selection techniques along with four optimization methods for prostate cancer classification is shown in Table 11.

A. COMPUTATIONAL COMPLEXITY OF THE METHODS

The performance of classifiers may also be analyzed through calculating the computational complexity of the algorithms. Based on the size of input (n), the computational complexity is calculated. The computational complexity reaches its least value when it is equivalent to O(1). The computational complexity will be increased surely as the number of input increases.

However, in this scenario, there is no dependence between complexity and input size, and this is one of the most desired entity for any kind of algorithm. If the computational complexity increases log(n) times with respect to increase in 'n', then it is denoted as O(logn). In this paper all the classifiers are hybrid in nature which classifies the optimized

TABLE 11. Computation complexity for all classifiers among four gene selection techniques along with four optimization methods for Prostate cancer.

Classifiers	Gene Selection Techniques	Optimization Methods			
		MHBOA	MBOA	SSOA	WOA
ANN	Relief - F	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log 2n)$
	Fishers Score	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log 2n)$
	Information Gain	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log n)$	$O(n^6 \log 2n)$
	SNR	$O(2n^6 \log n)$	$O(2n^6 \log n)$	$O(2n^6 \log n)$	$O(2n^6 \log 2n)$
LDA	Relief - F	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	Fishers Score	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	Information Gain	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	SNR	$O(2n^3 \log n)$	$O(2n^3 \log n)$	$O(2n^3 \log n)$	$O(2n^3 \log 2n)$
NBC	Relief - F	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	Fishers Score	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	Information Gain	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log 2n)$
	SNR	$O(2n^3 \log n)$	$O(2n^3 \log n)$	$O(2n^3 \log n)$	$O(2n^3 \log 2n)$
SVM (with RBF Kernel)	Relief - F	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log 2n)$
	Fishers Score	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log 2n)$
	Information Gain	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log n)$	$O(2n^2 \log 2n)$
	SNR	$O(4n^2 \log n)$	$O(4n^2 \log n)$	$O(4n^2 \log n)$	$O(4n^2 \log 2n)$

gene selected features. Table 11 shows the Computation Complexity for all the classifiers among four gene selection methods with four different optimization methods. From Table 11, the SVM (with RBF kernel) Classifier attained the computational complexity of $O(2n^2 \log n)$ for three gene selection methods like Relief – F, Fishers Score, and Information Gain along with three optimization methods such as MHBOA, MBOA and SSOA. $O(2n^2 \log 2n)$ computational complexity is reached in the SVM (with RBF kernel) Classifier with three gene selection methods like Relief – F, Fishers Score, and Information Gain for WOA Optimization method. This is the lowest computational complexity attained in this paper. The SVM (with RBF kernel) Classifier with SNR gene selection for three optimization methods like MHBOA, MBOA and SSOA reached next level of computational complexity of $O(4n^2 \log n)$. NBC and LDA classifiers are at the mean computational complexity of $O(n^3 \log n)$, $O(n^3 \log 2n)$, $O(2n^3 \log n)$and $O(2n^3 \log 2n)$. ANN Classifier retained its the Highest computational Complexity of $O(n^6 \log n)$, $O(n^6 \log 2n)$, $O(2n^6 \log n)$ and $O(2n^6 \log 2n)$ for four gene selection techniques as well as four optimization methods. Since ANN is a supervised classifier and this leads to higher computational complexity which includes the training of various layers of the Classifier.

VI. CONCLUSION AND FUTURE WORK

Cancer has a high association with genes which is responsible for carrying the information of human heritage. An understanding to the mechanism of cancers have been provided by the completion of human genome sequencing. To monitor hundreds of thousands of gene expression levels, biological technology is utilized. It is very important to develop efficient and precise classification models to differentiate between tumor and normal samples. But classification of the gene expression data has a lot of hassles due to the curse of dimensionality problem and small sample size. In this work, a tri-level feature selection methodology was adopted and then classified with suitable classifiers. The third best classification accuracy of 98.96% is obtained when Relief- F test with MHBOA is classified with ANN, Fishers Score with SSOA when classified with ANN, SNR with SSOA when classified with LDA and IG with SSOA when classified with SVM-RBF. The second-best classification accuracy of 99.22% is obtained when Relief-F test with MBOA is classified with Naïve Bayesian Classifier (NBC). The best classification results show when SNR with WOA is classified with Artificial Neural Network (ANN) giving a classification accuracy of 99.48%. Future works is to work with plenty of optimization techniques combined with advanced machine learning techniques for the efficient classification of prostate cancer.

REFERENCES

[1] A. R. Zlotta, M. Remzi, P. B. Snow, C. C. Schulman, M. Marberger, and B. Djavan, "An artificial neural network for prostate cancer staging when serum prostate specific antigen is 10 NG./ML. or less," *J. Urol.*, vol. 169, no. 5, pp. 1724–1728, May 2003.

[2] M. Remzi, T. Anagnostou, V. Ravery, A. Zlotta, C. Stephan, M. Marberger, and B. Djavan, "An artificial neural network to predict the outcome of repeat prostate biopsies," *Urology*, vol. 62, no. 3, pp. 456–460, Sep. 2003.

[3] P. Finne, R. Finne, A. Auvinen, H. Juusela, J. Aro, L. Määttänen, M. Hakama, S. Rannikko, T. L. J. Tammela, and U.-H. Stenman, "Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network," *Urology*, vol. 56, no. 3, pp. 418–422, Sep. 2000.

[4] C. Stephan, H. Cammann, A. Semjonow, E. P. Diamandis, L. F. Wymenga, M. Lein, P. Sinha, S. A. Loening, and K. Jung, "Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies," *Clin. Chem.*, vol. 48, no. 8, pp. 1279–1287, Aug. 2002.

[5] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub, "Multiclass cancer diagnosis using tumor gene expression signatures," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 26, pp. 15149–15154, Dec. 2001.

[6] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses," *Proc. Nat. Acad. Sci. USA*, vol. 98, no. 24, pp. 13790–13795, Nov. 2001.

[7] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring," *Science*, vol. 286, no. 5439, pp. 531–537, Oct. 1999.

[8] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, Jun. 2015, Art. no. 198363.
[9] Y. Wang and F. Makedon, "Application of relief-F feature filtering algorithm to selecting informative genes for cancer classification using microarray data," in *Proc. IEEE Comput. Syst. Bioinf. Conf. (CSB)*, Aug. 2004, pp. 497–498.
[10] N. Revathy and R. Balasubramanian, "GA-SVM wrapper approach for gene ranking and classification using expressions of very few genes," *J. Theor. Appl. Inf. Technol.*, vol. 40, no. 2, pp. 113–119, 2012.
[11] L. Sheng, R. Pique-Regi, S. Asgharzadeh, and A. Ortega, "Microarray classification using block diagonal linear discriminant analysis with embedded feature selection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1757–1760.
[12] K. L. Tang, W.-J. Yao, T.-H. Li, Y. X. Li, and Z. W. Cao, "Cancer classification from the gene expression profiles by discriminant Kernel-PLS," *J. Bioinf. Comput. Biol.*, vol. 8, no. 1, pp. 147–160, 2010.
[13] E. L. Markert, H. Mizuno, A. Vazquez, and A. J. Levine, "Molecular classification of prostate cancer using curated expression signatures," *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 52, pp. 21276-21281, 2011.
[14] Z. Jia, Y. Wang, and A. Sawyers, "Diagnosis of prostate cancer using differentially expressed genes in stroma," *Cancer Res.*, vol. 71, no. 7, pp. 2476–2487, 2011.
[15] M. Ram, A. Najafi, and M. T. Shakeri, "Classification and biomarker genes selection for cancer gene expression data using random forest," *Iranian J. Pathol.*, vol. 12, no. 4, pp. 339–347, 2017.
[16] S. Kim, "Spectral methods for cancer classification using microarray data," in *Proc. Int. Joint Conf. Comput. Sci. Optim.*, Apr. 2009, pp. 588–592, doi: 10.1109/CSO.2009.389.
[17] L. Xu, A. C. Tan, D. Q. Naiman, D. Geman, and R. L. Winslow, "Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data," *Bioinformatics*, vol. 21, no. 20, pp. 3905–3911, 2005.
[18] B. Zupan, J. Demšar, M. W. Kattan, J. R. Beck, and I. Bratko, "Machine learning for survival analysis: A case study on recurrence of prostate cancer," *Artif. Intell. Med.*, vol. 20, no. 1, pp. 59–75, Sep. 2000.
[19] B. Liu, Q. Cui, T.Jiang, S. Ma, "A combinational feature selection and ensemble neural network method for classification of gene expression data," *BMC Bioinf.*, vol. 5, p. 136, Sep. 2004, doi: 10.1186/1471-2105-5-136.
[20] S. M. Dhanasekaran, T. R. Barrette, and D. Ghosh, "Delineation of prognostic biomarkers in prostate cancer," *Nature*, vol. 412, no. 6849, pp. 822–826, 2001, doi: 10.1038/35090585.
[21] Z. Mao, W. Cai, X. Shao, "Selecting significant genes by randomization test for cancer classification using gene expression data," *J. Biomed. Inform.*, vol. 46, no. 4, pp. 594–601, Aug. 2013.
[22] O. Dangliyan, F. U. Yuksektepe, I. H. Kavakli, and M. Turkay, "Optimization based tumor classification from microarray gene expression data," *PLoS ONE*, vol. 6, no. 2, 2011, Art. no. e14579.
[23] S. H. Bouazza, K. Auhmani, and A. Zeroual, "Prostate cancer diagnosis based on Microarray gene expression profiles," *J. Eng. Technol.*, vol. 6, no. 2, pp. 282–291, Jul. 2018.
[24] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, and J. Manola, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, no. 2, pp. 9–203, 2002.
[25] Y. Liu, "Wavelet feature extraction for high-dimensional microarray data," *Neurocomputing*, vol. 72, nos. 4–6, pp. 985–990, Jan. 2009.
[26] Y. Peng, W. Li, and Y. Liu, "A hybrid approach for biomarker discovery from microarray gene expression data for cancer classification," *Cancer Inform.*, vol. 2, pp. 301–311, Jan. 2007.
[27] E. P. Xing, M. I. Jordan, and R. M. Karp, "Feature selection for high-dimensional genomic microarray data," in *Proc. ICML*, vol. 1, no. 2001, pp. 601–608.
[28] Z. Wang, "Neuro-fuzzy modeling for microarray cancer gene expression data," Oxford Univ. Comput. Lab., Oxford, U.K., Tech. Rep., 2005. [Online]. Available: https://www.cs.ox.ac.uk/people/zhenyu.wang/thesis.pdf
[29] H. S. Chang, "Converging marriage in honey-bees optimization and application to stochastic dynamic programming," *J. Global Optim.*, vol. 35, no. 3, pp. 423–441, Jul. 2006.
[30] L. Way Shen, H. Asmuni, and F. Cheng Weng, "A modified migrating bird optimization for university course timetabling problem," *Jurnal Teknologi*, vol. 72, no. 1, Dec. 2014.
[31] H. T. Ibrahim, W. J. Mazher, O. N. Ucan, and O. Bayat, "Feature selection using salp swarm algorithm for real biomedical datasets," *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 12, pp. 13–20, 2017.
[32] H. Zhang, H. Wang, N. Li, Y. Yu, Z. Su, and Y. Liu, "Time-optimal memetic whale optimization algorithm for hypersonic vehicle reentry trajectory optimization with no-fly zones," *Neural Comput. Appl.*, vol. 32, pp. 2735–2749, Oct. 2018.
[33] S. C. Ok and S. K. Sinha, "Construction equipment productivity estimation using artificial neural network model," *Construct. Manage. Econ.*, vol. 24, no. 10, pp. 1029–1044, Oct. 2006.
[34] M.-H. Lee, S. Fazli, and S.-W. Lee, "Optimal channel selection based on statistical analysis in high dimensional NIRS data," in *Proc. Int. Winter Workshop Brain-Comput. Interface (BCI)*, Feb. 2013, pp. 95–97.
[35] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.
[36] H. Byun and S.-W. Lee, "Applications of support vector machines for pattern recognition: A survey," in *Pattern Recognition With Support Vector Machines* (Lecture Notes in Computer Science), vol. 2388. Berlin, Germany: Springer, 2002, pp. 213–236.
[37] S. K. Prabhakar, H. Rajaguru, and S.-W. Lee, "A framework for schizophrenia EEG signal classification with nature inspired optimization algorithms," *IEEE Access*, vol. 8, pp. 39875–39897, 2020.

SUNIL KUMAR PRABHAKAR (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Anna University, Chennai, India, in 2012, 2014, and 2017, respectively. He is currently working as a Postdoctoral Research Fellow at the Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea. His research interests include signal processing, pattern recognition, and machine learning.

SEONG-WHAN LEE (Fellow, IEEE) received the B.S. degree in computer science and statistics from Seoul National University, Seoul, in 1984, and the M.S. and Ph.D. degrees in computer science from the Korea Advanced Institute of Science and Technology, Seoul, South Korea, in 1986 and 1989, respectively. He is currently the Hyundai-Kia Motor Chair Professor and the Head of the Department of Artificial Intelligence, Korea University. His research interests include pattern recognition, artificial intelligence, and brain engineering. He is a Fellow of IAPR and The Korean Academy of Science and Technology.