

Hybrid Bee Colony And Weighted Ranking Firefly Optimization For Cancer Detection From Gene Regulatory Sequences

M.Pyingkodi, Dr.S.Shanthi, M. Muthukumaran , K.Nanthini , K. Thenmozhi

Abstract: Gene regulatory sequence analysis is performed by subjecting a nucleic acid or peptide sequence to recognize its features, function, structure, or evolution. In bioinformatics, a regulatory sequence is a part of nucleic acid molecules which is capable for detecting the cancer in a living organism. In tumor cells, genes are not reacted normally to their regulators' activity and be the characteristics of the cancer disease. The several optimization techniques identify the disordered genes in given gene expression data. However an effective cancer detection model was not performed with less complexity. In order to improve the cancer detection with less complexity, Combinatorial Annealed Bee Colony and Weighted Ranking Firefly Optimization technique is developed. The main objective of proposed technique is to detect the cancer disease using gene sequences and information. Initially, number of training gene regulatory sequences is taken from the microarray dataset. After that, Annealed combinatorial bee colony optimization is performed to determine and select the gene regulatory sequences using population based search approach. Then, weighted ranking firefly optimization technique is applied to find the optimal gene sequences (i.e. firefly) based on their light intensity. This helps to identify the tumor and normal behavior of sequence by comparing the trained sequences with test gene sequences. As a result, deep learning of hybrid optimization techniques is used for detecting the cancerous tumor using these optimal gene sequences at an earlier stage with minimum time. Experimental evaluation of proposed technique is carried out with respect to number of gene sequences. The results confirmed that the proposed technique improves the performance results in terms of cancer detection rate, false positive rate, Time complexity and space complexity. Based on the results and discussions, proposed technique improved the cancer detection rate with minimum space as well as time complexity than the other methods.

Index Terms: Annealed Bee Colony ,cancer sequence ,Firefly optimization, Weighted Ranking Firefly Optimization
Gene Regulatory,

1. INTRODUCTION

In bioinformatics, gene sequence is constructed by nucleic acid or peptide sequence to recognize its structure alignment. The gene sequence information is essential for detecting the cancer at a previous stage. Genes controls the expression of other genes, which results in established structures as tumor or nontumor cells. Cancer involves due to the abnormal growth of cells and the gene expression deviates from its normal. Thus, selecting high appropriate gene sequences from gene expression data has become a significant field in bioinformatics data analysis. The several research works has been developed for detecting the cancer with gene sequences. Ant-colony-regulatory identification (ACRI) technique was introduced in [1] to detecting possible binding sites of transcription factor from the upstream of coexpressed genes. However, it failed to perform cancer detection with minimum time and space complexity. A genetic-algorithm-based mathematical approach was developed in [2] to determine optimum changes for preventing and detecting the cancerous tumors. But, an efficient optimization was not achieved to detect the cancerous tumor with less time. A circulating tumor cells (CTCs) was described in [3] for recognizing the most important genes used in breast cancer pathology. However, an optimal gene was not determined for cancer detection. mRNA- sequences of single Prostate Cancer Circulating Tumor Cells was presented in [4] for identifying over expressed genes. But, complexity during the gene identification remained unsolved.

Hepatocyte growth factor receptor (MET) network was developed in [5] during oral cancer detection. However the network was not provided the predictive value of new interactions. A new local and global preserving semi supervised dimensionality reduction based on random subspace algorithm (RSLGSSDR) was designed in [6] for diagnosis and treatment of cancers. But, performance analysis of space complexity remained unsolved. A penalized logistic regression model was developed in [7] for predicting a direction alteration of expressed genes by applying predicted transcription factor binding site. However an accurate prediction was not performed. A random Forest classification was introduced in [8] for predicting the new gastric cancer with similar genes based on selected features as well as prognostic genes were identified based on gene expressions. But, the false positive probability was not performed during the cancer prediction. In [9], a variety of expressed genes were involved in signaling pathways for gastric cancer detection. But, it failed to discover the optimal gene sequences for cancer detection. Sparse Probabilistic Principal Component Analysis (SPPCA) was developed in [10] to determine new gene-gene correlations relating with the cancer phenotypes. However, it failed to provide better prediction rules of cancer. The certain issues are observed from the above said literatures such as, high space and time complexity, failed to improve cancer prediction accuracy, lack of select optimal genes, and so on. To address such kind of issues, an efficient hybrid technique called Combinatorial Annealed Bee Colony and Weighted Ranking Firefly Optimization (CABC-WRFFO) technique is developed.

2. RELATED WORKS

A supervised learning hidden Markov model (HMMs) was developed in [11] for performing the cancer classification with the help of gene expression profiles. But, an optimal subset of genes was not selected for classification. CABC-WRFFO

- M.Pyingkodi ,Kongu Engineering College,India, pyingkodikongu@gmail.com
- Dr.S.Shanthi, Kongu Engineering College,India.
- Muthukumaran , Sunway University, Malaysia
- K.Nanthini , Kongu Engineering College,India.
- K. Thenmozhi, Selvam College of Technology, Namakkal, Salem, India.

technique effectively uses hybrid optimization technique for identifying the cancer at an earlier stage. In [12], a particle swarm optimization technique associated with decision tree classifier was developed to select an efficient gene from thousands of candidate genes in cancers detection. However, the hybrid search algorithm was not performed for reducing the execution time. CABC-WRFFO technique uses hybrid optimization technique for cancer detection with minimum time. A several machine learning (ML) techniques was presented in [13] to select certain genes for cancer prediction. However it failed to perform better statistical analysis for obtaining more accurate results for cancer detection. Therefore, a CABC-WRFFO technique performs efficient statistical analysis using hybrid optimization technique for cancer detection. A Hybrid search method Simulated Annealing with a Greedy Algorithm (SAGA) and Bayesian Networks was developed in [14] for detecting the transcriptional regulatory correlations among genes. However, the false discovery from true relationships was not improved. The proposed CABC-WRFFO technique reduces the false positive rate. A new gene expressions based colon classification approach (GECC) was introduced in [15] for performing the classification using colon gene samples and obtained two output classes as normal and malignant. However, an efficient technique was not used to assign weight for each gene. CABC-WRFFO technique assigned the weight to each gene sequence using weighted ranking firefly optimization technique. Support Vector Machine classifier was developed in [16] for cancer classification with gene expression data. But the optimization was not carried out for bio informatics data analysis. CABC-WRFFO technique performs optimization using Annealed combinatorial bee colony and weighted ranking firefly optimization. Binary particle swarm optimization (BPSO) technique was developed in [17] for gene selection on gene expression datasets. However, it failed to produce an optimal subset of genes from gene expression data. CABC-WRFFO technique selects an optimal gene sequences using weighted ranking firefly optimization technique. An evolutionary computation and multi objective optimization technique was introduced in [18] for solving the bioinformatics problem. However, cancer detection was not performed using optimized sequences. CABC-WRFFO technique performs cancer detection with the optimal gene sequences for reducing the complexity. Different gene expression pattern associated to breast cancer was detected in [19] using Cancer Genome Atlas dataset. However, the performance of breast cancer detection rate was not improved. Therefore, a CABC-WRFFO technique improves the performance analysis of cancer detection rate. cancer with gene expression data. But, optimal sequences were not selected for diseases prediction. Therefore, a CABC- WRFFO technique selects the optimal gene sequences from the training set to detect the cancer. A machine

3. PROPOSED METHOD

In microarray technology, large numbers of genes sequences are processed simultaneously to detect the cancer disease in an effective manner. In human body, the entire cells contain similar genetic sequences but these gene sequences may or may not be active. This differences in the activation of gene sequences helps to identify the function of the tumor cells in living organism. Microarray equipments help to identify several diseases namely various cancer disease, heart disease,

mental illness, infectious disease, and so on. But the cancer detection using gene sequences is important role in bioinformatics analysis. Therefore, the deep learning of hybrid optimization technique is developed to identify an optimal gene sequences for cancer detection.

3.1 HYBRID BEE COLONY AND WEIGHTED RANKING FIREFLY OPTIMIZATION TECHNIQUE

CABC-WRFFO technique contains a two processing steps for detecting the cancer from gene regulatory sequences. Gene regulatory sequences identification is also called as motif identification. The number of gene sequences is extracted from the microarray dataset. Then the Annealed combinatorial bee colony optimization (ACBCO) is applied to detect the gene sequences from the initial population. After that, weighted ranking firefly optimization technique (WRFFO) is used to find optimal gene sequences for cancer detection. The brief description of the CABC-WRFFO technique is explained in forthcoming subsections.

3.2 ALGORITHM

Input: Number of bees $b_1, b_2, b_3, \dots, b_n$

Output: Selected bees from initial population

Begin

Initialize the population of 'n' bees

For each bee ' b_i '

Perform local search to evaluate all possible constructive moves through the distance measure using (2)

Perform annealed selection to choose constructive forward move using (3)

Perform backward moves (i.e. All bees returned to nest)

Calculate the fitness function

Sort the bee population based on fitness function

Select bees from the sorted population

End for

End

3.3. ANNEALED COMBINATORIAL BEE COLONY OPTIMIZATION (ACBCO)

Annealed combinatorial bee colony optimization (ACBCO) is a population-based search algorithm to identify gene sequences for cancer detection. Combinatorial bee colony optimization technique is used for finding the sequences from a finite set of training set in microarray dataset. The ACBCO is inspired by the behavior of honey bees. In ACBCO, the bee's algorithm mimics the foraging strategy of honey bees (i.e. gene sequences) for determining the best solution. In ACBCO algorithm, food source represents gene sequence information and a population represents a colony of 'n' agents (i.e. honey bees) are used in search space. Each time small part of the bees constantly searches the environment for collecting the food sources. At first, the populations of 'n' bees are generated for obtaining the best solution. The population of 'n' bees is generated as follows,

$$b_i = b_1, b_2, b_3, \dots, b_n, \text{ where } i \in n(1)$$

From (1), b_i denotes number of bees (i.e. gene sequences) in search space. The bee searches neighborhood of the most promising solutions from the current path to perform forward constructive moves. During the searching process, total bees in the nest starts the constructive forward moves by performing the local search. Evaluate all the possible constructive moves of bees in search space. The proposed ACBCO technique uses annealed selection

approach to choose the constructive moves of bees for searching the food source. In ACBCO technique, selection probability of forward moves is measured based on distance of each individual. As the generation of population changes, distances changes and selection probability also changed. The proposed ACBCO technique computes Euclidean distance depending on the current position and next position.

$$Distance (D_i) = \sqrt{\sum_{i=1}^n (b_j - b_i)^2} \quad (2)$$

From (2) where b_j represent a bee in 'j'th position and b_i denotes bee in current position. After the distance measure, the annealed selection is applied to choose the next forward moves in search space with minimum distance. Therefore, the bee searches neighborhood position from the current position based on the minimum distances. This process is repeated for each bee to select the best moves and collects the food source. After that, the bees returned to nest and start backward moves. In general, bees perform a dancing ceremony which notifies other bees about the amount of food source they have collected, and the closeness of the neighborhood to the nest. Finally, the bees are sorted from the initial population based on fitness calculation. The fitness is used to select the quality of bees being suitable for identifying the tumor cells. Therefore, fitness of the bees is calculated based on their food sources information (i.e. sequence information) such as length of the sequences, similarity among the structure and binding sites. Based on fitness calculation, the bees are sorted and selected accordingly from the initial population for detecting the cancer. Annealed combinatorial bee colony optimization to select the gene sequences from the population for cancer detection through their sequence information. At first, the population of bees is initialized. After that, local search is performed to identify the neighborhood position for selecting the possible moves in the foraging strategy of honey bees. Followed by, ACBCO technique uses annealed selection approach to select best forward constructive move among the possible moves through the Euclidean distance measure and collecting the food source. After that, the bees are returned to nest by performing backward moves. Then the fitness of the honey bees is measured through the food source information. The bees are sorted based on their fitness calculation. Finally, the bees are selected from the sorted population. But an optimal gene sequences is not identified using Annealed combinatorial bee colony optimization for detecting the cancer disease. In order to find an optimal gene sequences, weighted ranking firefly optimization technique is developed.

3.4 WEIGHTED RANKING FIREFLY OPTIMIZATION TECHNIQUE

Optimal gene regulatory sequences are determined using weighted ranking firefly optimization technique. Weighted ranking firefly optimization is a metaheuristic approach and it activated by the flashing light behavior of fireflies. A metaheuristic is a partial search algorithm that provides an efficient solution in case of cancer detection. Therefore, CABCO-WRFFO technique uses WRFFO technique for identifying the optimal gene regulatory sequences from the population. WRFFO technique is designed based on flashing light approach of fireflies (i.e. gene regulatory sequences). In WRFFO technique, the group of fireflies moves towards the other fireflies by flashing light intensity with the objective function. WRFFO technique performed with following assumptions. The entire fireflies are unisexual. Any individual

firefly is attracted to other fireflies based on their light intensity since attractiveness of the firefly is directly proportional to their brightness. The less intensity firefly is attracted through the other brighter firefly. Therefore, the intensity of the fireflies is reduced as their distance increases. If there is no fireflies are brighter than particular firefly, it moves randomly to other and brightness of firefly gets increased.

The objective function of WRFFO technique is defined as an alignment of gene sequences (GS_A), gene sequence structure (i.e) binding sites (GS_{BS}), and certain chemical features of the gene sequences (GS_{CF}). Therefore, the objective functions is defined as follows,

$$f(x) = GS_A, GS_{BS}, GS_{CF} \quad (3)$$

From (3), $f(x)$ denotes an objective function of the each firefly. The WRFFO technique uses these objective functions for detecting the cancer in human body. The initial populations of fireflies (i.e. gene sequences) are generated as follows,

$$GS = GS_1, GS_2, GS_3, \dots, GS_n \in D \quad (4)$$

From (4), GS denotes an input of gene sequences obtained from ACBCO technique. Then the light intensity of all fireflies is formulated and it is directly proportional to the objective functions.

$$I(GS) \propto f(x) \quad (6)$$

From (5), $I(GS)$ denotes an intensity of gene sequences and $f(x)$ represents an objective functions. Then the firefly (GS_i) with higher light intensity attracts the other firefly (GS_j) to distinguish a benign tumor and malignant tumor. The malignant tumor is a mass of cells that lacks the ability to attack neighboring tissue based on the intensity level whereas benign tumor does not affects the neighboring tissue of the body. Therefore, the attraction of the fireflies are described as,

$$I(GS_j) > I(GS_i) \text{ Where } j = 1, 2, 3, \dots, n \text{ but } i \neq j \quad (6)$$

Therefore, the attractiveness of firefly's changes with the amount of absorption rate because their light intensity gets reduced based on the distance. The light intensity of the fireflies is a function of distance 'R' which is expressed as follows,

$$I(GS) = I_a \exp(-\gamma R) \quad (8)$$

From (7), $I(GS)$ denotes a light intensity of gene sequences (i.e. firefly) and I_a represents an actual light intensity and ' γ ' light absorption coefficient. The absorption coefficient of the light is based on the objects and also wavelength of light which is being absorbed. The low value of ' γ ' indicates a light of fireflies are poorly absorbed. From (7), 'R' denotes a distance between the gene sequences 'i' and 'j'. Therefore, the distance between the gene sequences are measured using Mahalanobis distance. Mahalanobis distance is a measure of the distance between gene sequences 'i' and 'j'. It is measured as follows,

$$R_{ij}^2 = (GS_i - GS_j)' * A^{-1} * (GS_i - GS_j) \quad (8)$$

From (8), R_{ij} denotes a Mahalanobis distance between sequences 'i' and 'j'. A^{-1} Represents an inverse covariance matrix whose sequences position is the covariance between ith and jth sequence of a random vector. 'T' denotes a transpose vector. Then the attractiveness of the gene sequences depends on the distance it is formularized as follows,

$$A_{ij} = R_{ij} \quad (9)$$

From (9), A_{ij} denotes an attractiveness of the gene sequences. If the gene sequence i move towards j with minimum distance. This helps to used for identifying all

possible binding sites from the expressed genes sequences for detecting the cancerous tumor. Based on the distance measure, the intensity for any pair of fireflies are updated as follows,

$$I^{t+1}(GS) = GS_i^t + A_{ij} e^{-\gamma R_{ij}^t} (GS_j^t - GS_i^t) + \rho_t \epsilon_t \quad (11)$$

From (10), $I^{t+1}(GS)$ represents an updated intensity of the gene sequences, $\rho_t \epsilon_t$ denotes a parameter controlling step size and ϵ_t represents a vector drawn from a Gaussian or other distribution. A_{ij} denotes an attractiveness of the gene sequences, R_{ij} distance between the fireflies and γ denotes a light absorption coefficient. Finally, all the sequences are ranked based on their updated light intensity. After updating, the light intensity of the firefly is ranked based on the weight value.

$$W_t \rightarrow \{GS_1, GS_2, GS_3, \dots, GS_n\} \quad (11)$$

From (11), W_t denotes a weight is applied to the entire fireflies. The higher weight is assigned to a firefly which has high light intensity than the other one. Based on the weight assignment, the fireflies are ranked in an ascending order. The high weighted fireflies are ranked first than the other which is expressed as follows,

$$r \rightarrow \{W_t(GS)\} \quad (12)$$

From (12) where 'r' denotes rank assigned to fireflies. As a result, WRFFO technique finds the globally optimized gene regulatory sequences to detect the cancer with minimum time. the weighted ranking firefly optimization algorithm for detecting the cancer with global best solution (i.e. gene regulatory sequences) to improve detection rate with minimum time. At first, the objective function is defined. Then the population of fireflies is initialized randomly and formulates their light intensity based on objective functions. Followed by, the distance between the pair of the fireflies is calculated using Mahalanobis distance measure. If the Light intensity of one firefly is higher than the other, then it moves and attracts to other firefly. After that, light intensity of all the fireflies is updated until termination condition is met. Finally, the weight is assigned to each fireflies depends on their light intensity. The weighted fireflies are ranked and selecting the fireflies with high rank. This helps to detect cancer using optimal gene sequences and improves the true positive rate with minimum time.

4. RESULT AND DISCUSSION

Experimental evaluation of Combinatorial Annealed Bee Colony and Weighted Ranking Firefly Optimization (ACBCO -WRFFO) technique is performed using MATLAB coding with colon tumor dataset. This dataset is taken from microarray dataset [21] and it contains 60 instances and 2,000 attributes. The attributes from 1 to 1999 denotes a genes. The final attributes 2000 is a class attributes which provides the two class labels normal and tumor. The normal represents the gene sequence in normal conditions and there is no abnormality is presented. The tumor class label represents an abnormal gene sequence (i.e. tumor). During the experimental evaluation, 1999 attributes are given as input and obtained the class label as output to identify the cancer. In training phase, optimal gene regulatory sequences are selected from training set by performing the deep learning of hybrid optimization and it compared with given test set to detect the normal and tumor affected sequences. The evaluation of ACBCO -WRFFO technique is compared with existing ant-colony-regulatory-identification (ACRI) [1] and genetic-algorithm-based optimization [2]. The experiment is conducted on the factors

such as cancer detection rate, false positive rate, time complexity and space complexity. Result and discussion of ACBCO -WRFFO technique is described and compared with existing ACRI [1] and genetic-algorithm-based optimization [2]. The performance analysis is carried out with different parameters such as cancer detection rate, false positive rate, time complexity and space complexity with respect to number of gene sequences. The results are discussed with the help of tables and graph values.

4.1 IMPACT OF CANCER DETECTION RATE

Cancer detection rate is defined as the ratio of number of optimal gene regulatory sequences are determined to the total number of sequences in micro array dataset. Cancer detection rate is measured as, $CDR = \frac{(n - \text{No. of optimal gene sequences for cancer detection})}{n} * 100$

Where 'CDR' denotes a cancer detection rate and 'n' denotes total number of gene sequences and it is measured in terms of percentage (%).

Table 1 Tabulation for Cancer detection rate

NO. OF GENE SEQUENCES	CANCER DETECTION RATE (%)		
	ACBCO - WRFFO	ACRI	GENETIC-ALGORITHM BASED OPTIMIZATION
100	85	61	72
200	86	65	74
300	87	68	75
400	88	70	77
500	89	72	80
600	90	75	82
700	91	77	83
800	92	79	84
900	93	80	86
1000	95	81	87

Table 1 shows the performance comparison results of cancer detection rate versus number of gene sequences. Performance of cancer detection rate is performed with three different methods namely CABCO-WRFFO technique and existing ACRI [1] and genetic-algorithm-based optimization [2]. As shown in table, cancer detection rate is improved using ACBCO -WRFFO technique when compared to existing methods.

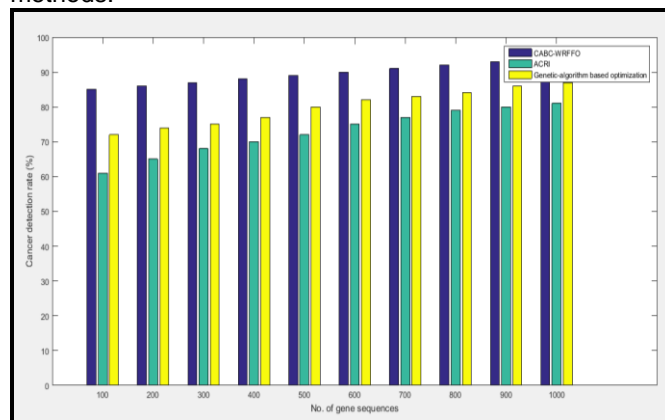


Figure 4 performance result of cancer detection rate

which compared to existing methods. This is because, the CABCO-WRFFO technique performs deep learning about the hybrid optimization technique for

determining the optimal sequences for detecting the cancer at an earlier stage. At first, Combinatorial Annealed Bee Colony (ACBCO) optimization technique is applied to select the number of gene sequences from the micro array dataset. The populations of bees are initialized and perform local search to identify the neighborhood position. Based on local search, the next forward moves of honey bees are selected for searching the food source. The fitness of each honey bees are measured based on food source information. Then the bees are sorted and selected from the initial population. In order to determine the optimal gene sequences, CABCO-WRFFO technique uses weighted ranking firefly optimization. The population of the fireflies is initialized randomly and measure the light intensity based on objective functions. The firefly with less light intensity is attracted by higher one. Followed by, the light intensity of all the fireflies is updated. The weight value of all fireflies is assigned about their updated light intensity. Finally, the fireflies are ranked based on the weight value. The high ranked fireflies are selected to detect cancer or normal. This helps to detect the cancerous tumor with optimal gene regulatory sequences. As a result, cancer detection rate is increased by 24% and 12% when compared to existing ACRI [1] and genetic-algorithm-based optimization [2] respectively.

4.2 FALSE POSITIVE RATE

False positive rate is measured as the ratio of number of optimal gene regulatory sequences are incorrectly selected to the total number of sequences in microarray dataset.

Table 2 Tabulation for false positive rate

NO. OF GENE SEQUENCES	FALSE POSITIVE RATE (%)		
	ACBCO - WRFFO	ACRI	GENETIC-ALGORITHM-BASED OPTIMIZATION
100	21	43	32
200	24	45	34
300	26	46	36
400	28	48	38
500	32	51	40
600	34	53	41
700	38	54	44
800	40	56	46
900	44	57	49
1000	47	58	53

Table 2 describes the performance results of false positive rate versus number of gene sequences in microarray dataset. The performance is compared with three different methods ACBCO -WRFFO technique, ACRI [1] and genetic-algorithm-based optimization [2]. The false positive rate of proposed ACBCO -WRFFO technique is considerably reduced when compared to existing methods. It is clearly evident that the false positive rate is considerably reduced using ACBCO -WRFFO technique than the existing methods. Performance of false positive rate is reduced by applying hybrid optimization techniques. At first, the combinatorial annealed bee colony optimization is applied for selecting the gene sequence from the total population. ACBCO is a population based search approach to select the gene sequences for cancer detection. In ACBCO, the bee performs local search to detect forward moves for foraging strategy of food from the current position. From that calculation, the annealed selection is applied to choose the best moves to search food source and returned back to the nest with collected food source. As a result, the bees are sorted and selected based on sequence information.

Weighted ranking firefly optimization technique is applied to select an optimal gene sequences from the selected genes from ACBCO. This process is used to determine optimal gene sequences for detecting the cancerous tumor or normal. As a result, the false positive rate is considerably reduced by 36% and 20% when compared to existing ACRI [1] and genetic-algorithm-based optimization [2] respectively.

4.3 TIME COMPLEXITY

The result of time complexity is significantly reduced. This is because of weighted ranking firefly optimization is used to determine optimal sequence for cancer detection. The intensity of fireflies (i.e. gene sequences) is formulated based on objective functions. The objective function of WRFFO technique is defined as gene sequences alignment, structure of gene sequences and chemical features. Then the distance between the fireflies is measured using Mahalanobis distance measure. If low light intensity of firefly move towards the higher intensity firefly. Subsequently, the light intensity of the entire fireflies is updated. Then the weight is assigned to an each firefly based on their light intensity. Finally, the each firefly is ranked according to their weight value. The ranked sequences are selected as optimal sequences to detect the cancer with minimal time. As a result, the time complexity of CABCO-WRFFO technique is significantly reduced by 35% and 22% when compared to existing ACRI [1], genetic-algorithm-based optimization [2] respectively.

4.4 IMPACT OF SPACE COMPLEXITY

The deep learning of proposed hybrid CABCO-WRFFO technique effectively determines an optimal gene sequences from the microarray dataset. An abnormal growth of tissue is identified effectively using optimal gene sequences with less time and higher detection rate. Therefore, the storage capability of optimal gene sequences is reduced using CABCO-WRFFO technique. Let us consider 100 genes sequences as input, the space complexity of proposed CABCO-WRFFO technique is 51KB whereas 84KB and 73KB using existing ACRI [1], genetic-algorithm-based optimization [2] respectively. This shows the space complexity of CABCO-WRFFO is reduced by 29% and 22% when compared to existing ACRI [1] and genetic-algorithm-based optimization [2] respectively.

As a result, the above performance results show that the combinatorial annealed bee colony and weighted ranking firefly optimization technique improves the cancer detection rate with minimum space and time complexity.

Table 2 Tabulation for Space Complexity

NO. OF GENE SEQUENCES	CABCO-WRFFO	ACRI	GENETIC ALGORITHM
	100	51	84
200	55	90	79
300	63	94	82
400	67	98	90
500	72	106	95
600	78	119	105
700	95	128	113
800	103	137	127
900	118	143	133
1000	127	150	142

5. CONCLUSION AND FUTURE WORK

The deep learning of Combinatorial Annealed Bee Colony and Weighted Ranking Firefly Optimization (ACBCO -WRFFO) technique is presented to detect the cancer with minimum space time and complexity. ACBCO -WRFFO technique considers two processing steps for improving the cancer detection at an earlier stage using optimal gene sequences. At first, the population based search approach namely annealed combinatorial bee colony optimization (ACBCO) is used for selecting the gene sequences from training set based on fitness of the honey bees through the food source information. These sequences are given to the input of weighted ranking firefly optimization (WRFFO) for determining the optimal sequences for cancer detection. This is done by assigning the weighted rank to each firefly. Experimental evaluation of ACBCO -WRFFO technique is performed using colon tumor dataset with different parameters such as cancer detection rate, false positive rate, time complexity and space complexity. The results analysis of CABO-WRFFO technique improves cancer detection rate with minimum false positive rate, time and space complexity than the state-of-art methods.

REFERENCES

- [1] WeiLiu, HanwuChen, LingChen, "An ant colony optimization based algorithm for identifying gene regulatory elements", *Computers in Biology and Medicine*, Elsevier, Volume 43, 2013, Pages 922–932
- [2] Danh Cong Nguyen and Farhad Azadivar, "Application of Computer Simulation and Genetic Algorithms to Gene Interactive Rules for Early Detection and Prevention of Cancer", *IEEE Systems Journal*, Volume 8, Issue 3, 2014, Pages 1005 - 1013
- [3] Stelios Sfakianakis , Ekaterini S. Bei , Michalis Zervakis, Despoina Vassou, Dimitrios Kafetzopoulos, "On the Identification of Circulating Tumor Cells in Breast Cancer", *IEEE Journal of Biomedical and Health Informatics* , Volume 18, Issue 3, 2014, Pages 773 - 782
- [4] Gordon M. Cann, Zulfiqar G. Gulzar, Samantha Cooper, Robin Li, Shujun Luo, Mai Tat, Sarah Stuart, Gary Schroth, Sandhya Srinivas, Mostafa Ronaghi, James D. Brooks, AmirAli H. Talasaz, "mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer" *PLoS One*, Volume 7, Issue 11, 2012, Pages 1-10
- [5] K. Kalantzaki , E. S. Bei, K. P. Exarchos ,M. Zervakis , M. Garofalakis ,D. I. Fotiadis, "Nonparametric Network Design and Analysis of Disease Genes in Oral Cancer Progression", *IEEE Journal of Biomedical and Health Informatics* , Volume 18, Issue 2, 2014, Pages 562 – 573
- [6] Xianfa Cai , Jia Wei , Guihua Wen ,Zhiwen Yu, "Local and Global Preserving Semisupervised Dimensionality Reduction Based on Random Subspace for Cancer Classification", *IEEE Journal of Biomedical and Health Informatics*, Volume 18, Issue 2, 2014, Pages 500 – 507
- [7] Hong Hu, Jingting Xu, Yang Dai, "Regulatory Elements in Low-Methylated Regions Predict Directional Change of Gene Expression", *IEEE Journal of Biomedical and Health Informatics*, Volume 19, Issue 4, 2015, Pages 1293 - 1300
- [8] Xuesong Wu, Haoran Tang, Aoran Guan, Feng Sun, HuiWangc, Jie Shua, "Finding gastric cancer related genes and clinical biomarkers for detection based on gene–gene interaction network", *Mathematical Biosciences*, Elsevier, Volume 276, 2016, Pages 1-7
- [9] Hao Li, Beiqin Yu, Jianfang Li, Liping Su, Min Yan, Jun Zhang, Chen Li, Zhenggang Zhu , Bingya Li, "Characterization of Differentially Expressed Genes Involved in Pathways Associated with Gastric Cancer", *PLoS one*, Volume 10, Issue 4, 2015, Pages 1-17
- [10] Shuichi Kawano, Teppei Shimamura, Atsushi Niida, Seiya Imoto, Rui Yamaguchi, Masao Nagasaki, Ryo Yoshida, Cristin Print, and Satoru Miyano, "Identifying Gene Pathways Associated with Cancer Characteristics via Sparse Statistical Methods", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Volume 9, Issue 4, 2012 , Pages 966 - 972
- [11] Thanh Nguyen, Abbas Khosravi, Douglas Creighton, Saeid Nahavandi, "Hidden Markov models for cancer classification using gene expression profiles", *Information Sciences*, Elsevier, Volume 316, 2015, Pages 293-307
- [12] Kun-Huang Chen, Kung-Jeng Wang, Min-Lung Tsai, Kung-Min Wang, Angelia Melani Adrian, Wei-Chung Cheng, Tzu-Sen Yang, Nai-Chia Teng, Kuo-Pin Tan and Ku-Shang Chang, "Gene selection for cancer identification: a decision tree model empowered by particle swarm optimization algorithm", *BMC Bioinformatics*, Volume 15, Issue 49, 2014, Pages 1-10
- [13] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzisc, Dimitrios I. Fotiadis, "Machine learning applications in cancer prognosis and prediction", *Computational and Structural Biotechnology Journal*, Elsevier, Volume 13, 2015, Pages 8-17
- [14] Emmanuel S. Adabor, George K. Acquaaah-Mensah , Francis T. Oduro, "SAGA: A hybrid search algorithm for Bayesian Network structure learning of transcriptional regulatory networks", *Journal of Biomedical Informatics*, Elsevier, Volume 53, 2015, Pages 27–35
- [15] Saima Rathore ,Mutawarra Hussain, Asifullah Khan, "GECC: Gene Expression Based Ensemble Classification of Colon Samples", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Volume 11, Issue 6, 2014, Pages 1131 – 1145
- [16] Heena Farooq Bhat, "Evaluating SVM Algorithms for Bioinformatics Gene Expression Analysis", *International Journal of Computer Science Engineering (IJCSSE)*, Volume 6, Issue 02, 2017, Pages 42-52
- [17] Mohd Saberi Mohamad , Sigeru Omatu , Safaai Deris , Michifumi Yoshioka, "A Modified Binary Particle Swarm Optimization for Selecting the Small Subset of Informative Genes From Gene Expression Data", *IEEE Transactions on Information Technology in Biomedicine*, Volume 15, Issue 6, 2011, Pages 813 – 822
- [18] Álvaro Rubio-Largo, Miguel A. Vega-Rodríguez ,David L. González-Álvarez, "A Hybrid Multiobjective Memetic Metaheuristic for Multiple Sequence Alignment", *IEEE Transactions on Evolutionary Computation*, Volume 20, Issue 4, 2016, Pages 499 – 514
- [19] Zhenzhen Huang, Huilong Duan, and Haomin Li, "Identification of Gene Expression Pattern Related to Breast Cancer Survival Using Integrated TCGA Datasets and Genomic Tools", *BioMed Research International*, Hindawi Publishing Corporation Volume 2015, September 2015, Pages 1-10
- [20] Shaurya Jauhari and S.A.M. Rizvi, "Mining Gene Expression Data Focusing Cancer Therapeutics: A

Digest", IEEE/ACM Transactions on Computational
Biology and Bioinformatics, Volume 11, Issue 3, 2014,
Pages 533 - 547

- [21] Colon tumor dataset:
<http://csse.szu.edu.cn/staff/zhuzx/Datasets.html>