ADAPTIVE FUZZY WITH ENHANCED GENETIC ALGORITHM TO DIAGNOSING BREAST CANCER

K. Thenmozhi, K. Kanimozhi, K. Ramesh, M.Pyingkodi*

Abstract

Cancer is one of the significant biological diseases in the real world and the foremost cause of death worldwide. Uncontrolled growth of cells in a human body causes cancer. Especially breast cancer is the most common cancer in women. Identifies a specific cancer is a crucial task in the overgrowth of biological information. Most of the researchers challenged to diagnose, but no one can unerringly prevent this dangerous disease. Generally, the machine learning technique, genetic methods used to diagnosis the diseases. In existing, the hierarchical clustering used to diagnosis a breast cancer which is suitable for the considerable volume of data set but challenging to identify the mutant data, hard to achieve accuracy, specificity, Sensitivity, and f-measures.

Meanwhile, the genetic algorithm is also well apt to identify this disease, and only the issue is that it has enhanced the optimal local search; it lost the global optima search. To enhance both search to get optimum results. It introduced a novel Adaptive Fuzzy method with an Enhanced Genetic Algorithm (AFEGA) to diagnose breast cancer in DNA sequence data set. Cuckoo search enhances the global search, and genetic algorithm enhances the local search to import cuckoo concept to genetic. Data granule was done by adaptive fuzzy. These AFEGA achieve an accurate breast cancer detection rate, improves the specificity, Sensitivity, and f-measures.

Keywords: Breast cancer diagnosis; Adaptive fuzzy; DNA sequence; Genetic algorithm; Cuckoo search;

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

I. INTRODUCTION

Cancer is one of the topmost causes of disease in the world [1]. Uncontrolled growth of cells in a body makes a mutation that is said to cancer. A GA is a searching algorithm based on the procedure of natural genetics. It is based on an iterative process that includes a constant-size population of individuals. A finite string of symbols characterizes every individual, and a probable solution is encoded in a specified problem space. This space is termed a search space that includes all the probable solutions to resolve the specified problem. GA is applied to spaces that are huge to be comprehensively searched.

GA is a local search approach used for identifying the suitable solution to the specified issue and search optimization. GA is a specific evolutionary class that employs evolutionary biology mutation, crossover, inheritance, and selection methods. It is constructed as a computer simulation in which several abstract characterizations of candidate solutions are developed for better solutions. In every generation, the fitness of the entire population assessed by individuals is chosen from the current population and customized to produce a new population. It is a necessity to improve the global search for the optimum solution.

The CS algorithm is one of the essential optimization algorithms to deal with the optimization problem [2] to enhance the global search optima. The cuckoo bird lays their eggs in the nest of the host bird. To overcome the GA's problem, the GA with CS is proposed to cluster the DNA data. Crossover, fitness in genetic value will be calculated by cuckoo's eggs. To identify the overlap eggs in the nest using

soft computing techniques such as adaptive fuzzy.

To make more effective and better performance, the FL has to be adaptive. The adaptive fuzzy [3] minimizes the objective function and gives the optimum cluster. [4] described the adaptive fuzzy that is a fuzzy controller applied with a training method which adjusts the parameters of fuzzy system with respect to numerical input-output data. It is mainly used for the nonlinear structure. It continuously produces the rules for fuzzy techniques.

II. LITERATURE SURVEY

The medical field's main issue is disease diagnosis; disease prediction is difficult for the huge volume of data. This feature selection using genetic methods combines three various classifiers such as ANN (Artificial Neural Network, PS-classifiers, GA based classifiers used to detect the breast cancer disease using Wisconsin Breast Cancer (WBC) data set. However, the problem of this work is if it increases the Sensitivity, the specificity was decreased and vice versa [5]. This article highlighted the main issues of cancer diagnosis and drug diagnosis, which is used to classify cancer using a hybrid meta-heuristic algorithm (GALA). It reduces the complexity, but it is not suitable for all the data set. It is suitable only for the selective data set and also lack sensitivity identification in cancer diagnosis [6].

Distribution-based Fuzzy Estimate Spectral Clustering (DFESC) technique used to diagnosis cancer for DNA sequence dataset. Based on the membership value it identifies the overlapping among the centroid data set, which improves the accuracy of detection and reduce the time complexity and the lacking in feature selection for more accuracy [7]. Improved Cuckoo Search Algorithm (ICSA) also used to diagnosis the cancer disease with DNA sequence data set which is taken from UCI repository [8,9].

The breast cancer diagnosis is based on classifiers Support Vector Machine (SVM) and Particular Swarm Optimization (PSO) techniques. It referred the dataset as Wisconsin Breast Cancer Dataset (WBCD) for detects the breast cancer in accurate manner. It was suitable only for tiny volume of data. It takes much time and space for huge volume of data set. The main drawback of PSO how low convergence, easily descend into local optimum value for high dimensional data set. To overcome these issues by proposed techniques, which is used distributed cluster for handling huge volume of data set. Meanwhile, the GA and CS combined with adopting the local and global minima and maintaining the convergence rate for each iteration [10,11].

III. ADAPTIVE FUZZY WITH ENHANCED GENETIC ALGORITHM

A cell is the central core of all living organisms. The cell contains a chromosome, which contains DNA. It translates to RNA, which transcript to the DNA sequence. The overgrowth of cells in the human body causes cancer. Cell mutated in the breast called breast cancer. Types of cancer depend upon the parts of the body where the growth is uncontrolled.

The base pair of DNAs is A, G, T, and C. One pair was mutated to some other pair. The DNA sequences as input which is getting from the DNA data set for breast cancer analysis. The abnormal DNA sequences from the DNA data set are considered as cancerous DNA sequences (cancer data), i.e., static data, and the standard DNA sequences from the DNA data set are considered as active data.

During this cancer diagnosis, the active or inactive similar data are clustered from the large volume of DNA data set with the performance of distributed clustering. Hence the heuristic optimization as AFEGC is introduced in a distributed computing environment to cluster large numbers of DNA sequences for performing cancer detection. During

the distributed clustering, the proposed AFEGC successfully allocates the computing resources in an effective manner.

Besides, the diversity process of global minima is accelerated by modifying the GA algorithm with a CS. Additionally, the soft computing technique termed as AFL is introduced for enhancing the clustering accuracy. This is calculating the membership value of each DNA sequence and identified the overlapping among the DNA sequence and centroid. Then the AFEGC is implemented for performing distributed clustering in the breast cancer disease diagnosis. The figure illustrates the architecture of AFEGC.

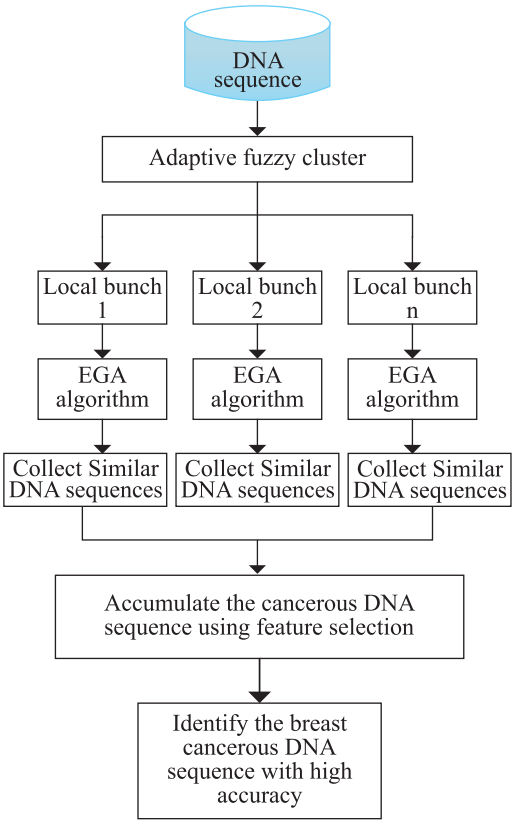

Figure .1 Architecture of AFEGA

DNA sequence locally granule into the separate cluster using AFL. Then, the updated genetic algorithm was applied for each granule to collect similar DNA sequences, which makes breast cancer based on feature selection. Finally, diagnosis the breast cancer with the combination of distributed clustering, adaptive fuzzy logic, updated genetics, and feature selection termed AFEGA.

```
Begin
    Make an opening population of n host nests
    Apply adaptive fuzzy to granule
    Assess its fitness F
    if { EMBED Equation.DSMT4}
      Replace { QUOTE{ QUOTE j}} by the new solution { QUOTE{ QUOTE i}}
      end if
    selection done with Cuckoo search
    while ( t < MG)
        for every nest
            Find a cuckoo type randomly (say, { EMBED Equation.DSMT4}
            { QUOTE{ QUOTE i}}
            Check the type of the cuckoo based on feature
            selection
            if type = common cuckoo
                    two eggs with the help of crossover
                    operator and
                    choose best one
            else if type = other cukcoo
                    two eggs using crossover with
                    uniform mutation
                    operator and choose the best one
             else
                    Create an egg with random solution
            end if
                    Choose an egg with the
                    Nastiest solution in
            Rank the eggs based on the solution
        end for
    end while
```

End

Figure 2 Pseudo code for AFEGA

IV. RESULTS AND DISCUSSION

The AFEGA is implemented by using MATLAB. It utilizes DNA dataset as BRCA2. The main aim of this data is to detect breast cancers and non-cancer. During the experiment, the AFEGA is compared with the existing methods like PSO with SVM. A mutation can change in genetic sequence.

Original individual nucleotide sequence:

TGCCTAAGTATT

TGCCAAAAGTAT

TGCCAAAAGTAT

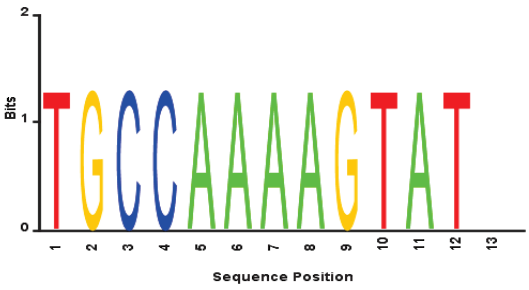

Figure 3. Original sequence motif

Mutated individual nucleotide sequence:

TGCTTAAGTAGT

AGCATAGTATAT

AGCCAAAAGTAT

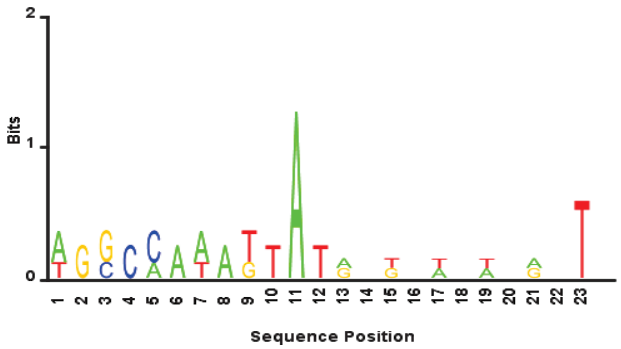

Figure 4. Cancerous sequence motif

Normal DNA sequence: ATGC AGGC ATCA ATTC

ATGC AGGC ATCA ATTC

ATGC AGGC ATCA ATTC

ATGC AGGC ATCA ATTC

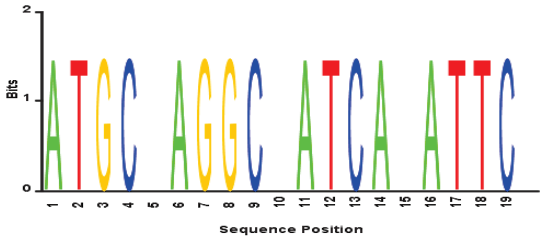

Figure 5. DNA sequence motif

DNA mismatch: ATGC AGGC ATCA ATTC

ATGC TCGG ATCA ATTC

ATGC AGGC ATAA ATTC

ATAC AGTC ATCA ATTC

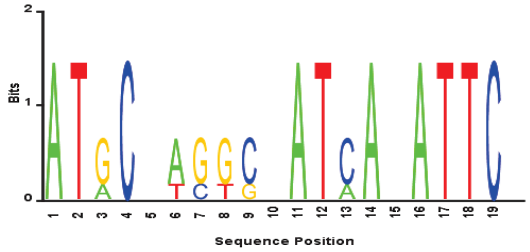

Figure 6. Mutated DNA sequence motif

A. Measurement of Specificity

The specificity measures the quality. The ratio of corrected discovers negative sequence to how much positive sequence.

Specificity= TN/TP+TN

Number of DNA sequences	Specificity (%)	
	PSO - SVM	AFEGA
500	54	66
1000	58	71
1500	60	78
2000	63	84
2500	67	90

Table 1 Specificity for PSO-SVM and AFEGA

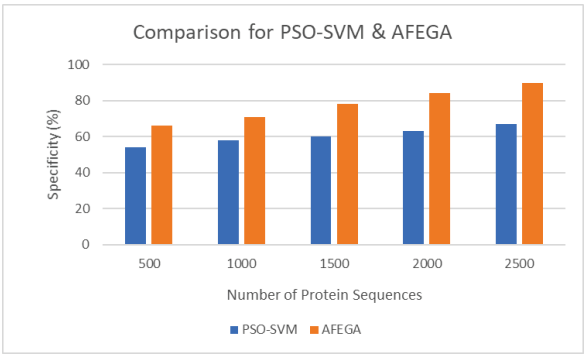

Figure 7 Comparisons for PSO-SVM and AFEGA in terms of specificity

B. Measurement of Sensitivity

The Sensitivity measures the quality. The ratio of corrected discovers positive sequence to how much positive sequence.

Sensitivity =TP/FN+TP

Number of DNA sequences	Specificity (%)	
	PSO - SVM	AFEGA
500	65	75
1000	71	79
1500	73	83
2000	78	86
2500	81	92

Table 2 Specificity for PSO-SVM and AFEGA

Figure 8 Comparisons for PSO-SVM and AFEGA in terms of Sensitivity

C. Result Analysis of F-measure

The F-measure is the quantity and quality of accuracy. F-measure was calculated based on specificity and Sensitivity F-measure= 2*(specificity *sensitivity) / (specificity + sensitivity)

Number of DNA sequences	F-measure (%)	
	PSO - SVM	AFEGA
500	58.99	70.21
1000	63.84	74.79
1500	65.86	80.42
2000	69.70	84.99
2500	73.34	90.99

Table 3 F-measure for PSO-SVM and AFEGA

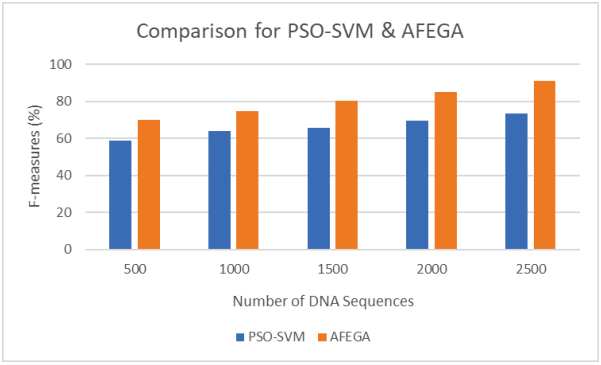

Figure 9 Comparisons for PSO-SVM and AFEGA in terms of F-measure

V. CONCLUSION

An optimum solution, searches are done based on local and global. However, GA enhances the local search. CS enhances the global search. GA & CS interpreted to get optimum results. The motive for GA usage to determine the exploration space more efficiently than the other methods is clarified, and it is confirmed why GA-related techniques produce accuracy. The performance of clusters with GA, CS, and AFCM is explained by the results, which show the boosted performance of transformation. This helps to enhance the local and global optimum values. AFEGA determines the cancerous DNA sequences among a large number of DNA sequences. The aim is to enhance the accuracy of cancer detection, improves specificity and Sensitivity. Specificity and Sensitivity are directly proposed to F-measure. The proposed AFEGA shows a better performance than existing methods in terms of accuracy, specificity, Sensitivity, and f-measures.

REFERENCES

[1] Zaorsky, NG, Churilla, TM, Egleston, BL, Fisher, SG, Ridge, JA, Horwitz, EM, & Meyer JE 2017, 'Causes of death among cancer patients', Annals of Oncology, vol. 28, no. 2, pp. 400-407.

[2] Yang XS & Deb, S 2009, 'Cuckoo search via lévy flights', In: Proceedings of world congress on Nature & Biologically Inspired Computing (NaBIC 2009, India), pp. 210-214.

[3] Pham, DL & Prince, JL 1999, 'An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities', Pattern Recognition Letters vol. 20, no. 1, pp. 57-68.

[4] Zuo, WL, Wang, ZY, Liu, T & Chen, H-L 2013, 'Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach', Biomedical Signal

Processing and Control,vol.8,no.4,pp. 364-373.

[5] Shokoufeh A, Hadi S, Alireza R and Saeid E 2016, 'Feature selection using genetic algorithm for breast cancer diagnosis: an experiment on three different datasets', Iran J Basic Med Sci. Vol.19, No.5, pp. 476–482.

[6] Habib M, Ali N, Balal S, Ali M-N 2017, 'A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata',Informatics and medicine unlocked , Vol. 9,pp. 246-254.

[7] Thenmozhi, K, Karthikeyani Visalakshi, N & Shanthi, S, 2018, 'Distribution based Fuzzy Estimate Spectral Clustering for Cancer Detection with Protein Sequence and Structural Motifs', Asian Pacific Journal of Cancer Prevention, vol. 19, no. 7, pp. 1935-1940.

[8] Thenmozhi, K, Karthikeyani Visalakshi, N, Shanthi, S, & Pyingkodi, M, 2018, 'Distributed ICSA Clustering Approach for Large Scale Protein Sequences and Cancer Diagnosis', Asian Pacific Journal of Cancer Prevention, vol. 19, no. 11, pp. 3105-3109.

[9] Ibrahim O 2018, 'Diagnosis of Breast Cancer Using Power Swarm Optimization with SVM',Cyprus International University.

[10]Lukasz C & Agnieszka D 2019, 'Efficient Genetic Algorithm for Breast Cancer Diagnosis', In book: Information Technology in Biomedicine (pp.64-76).

[11] Tanzeem K, Amrit S, Sadhna K,M 2014, 'Application of Genetic Algorithm for Cancer Diagnosis by Feature Selection', International Journal of Engineering Research & Technology (IJERT), Vol. 3, No. 8.