International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 8
August 2017

Breast Cancer Risk Evaluation By Firefly Optimization Algorithm

K. Saravana Kumar[1], Arthanariee A. M.[2]

[1] Research and Development Centre, Bharathiar University, Coimbatore ,Tamil Nadu, India.
Associate Professor, Christ University, Bangalore, Karnataka,India.

[2] Dean (Retd.), Science and Humanities, Nehru Institute of Technology, Kaliyapalayam, Coimbatore, India.

Multiple factors are influencing to get the breast cancer. It is rapidly increasing in the developed countries and developing countries. There are multiple reasons for this deadly disease. There are multiple algorithms to precisely find out the malignant tissue but the accuracy factor is very important to diagnose the disease. Normally the data base automatically updated when the patient approach the hospital for the purpose of diagnosis. These values can be compared and analyzed with medical images. It has to be amalgamated as single schema for the further comparison to bring out the better accuracy in the result. The optimization techniques are vastly used in different sectors because of its simplicity, robustness and efficiency to solve sophisticated optimization problems and all these features are the part of firefly Optimization algorithm (FA or FFA). The algorithm is based on the flashing behavior of fireflies which is compared with how the cancerous cell attracted to other cells in the human body. In the algorithm, the randomly generated answers will be considered as fireflies, and brightness is allocated based on their performance of the objective function.

Keywords — Influencing, multiple algorithms, accuracy factor, diagnosis, single schema, amalgamated, optimization algorithm, brightness, objective function

I. INTRODUCTION

Breast Cancer is the most prevalent cancer and major terminal illness among women worldwide. It usually occurs in the fibro glandular area of breast tissue. Every year it is increasing in developed and developing countries [1]. In India, we are now witnessing more number of patients being diagnosed with breast cancer to be in the younger age groups (in their thirties and forties). Twenty five years back, out of every 100 breast cancer patients, 2% were in 20 to 30 years age group, 7% were in 30 to 40 and so on. 69% of the patients were above 50 years of age but the present statistics gives 4% are in 20 to 30 yrs age group, 16% are in 30 to 40, 28% are in 40 to 50 age group. This gives an alarming picture almost 48% patients are below 50. The number of patients increasing in the age group of 25 to 40, and the trend is very disturbing one. The reason for this every year it is adding higher numbers of younger patients in our population pyramid. This denotes broad at the base and middle and narrow at the top, which gives meaning that huge population in the younger age group and much lesser in the older age group [2].

The developed country like US about 1 in 8 women (about 12%) will develop invasive breast cancer over the course of their lifetime. In particular this year 2017, an estimated 252,710 new cases of invasive breast cancer are expected to be diagnosed in women. The estimation says a man's lifetime risk of breast cancer is about 1 in 1,000.

The previous two decades the breast cancer was increasing but after the year of 2000 the trend has been changed. It dropped by 7% from 2002 to 2003 alone because of awareness and reduced use of Hormone Replacement Therapy (HRT) by women. The study shows that the close connection between HRT and increased breast cancer risk. Breast cancer also linked to gene mutations nothing but abnormal changes inherited from the parents. There are two most common genes are BRCA1 and BRCA2 and it suggest that women with a BRCA1 mutation have a 55-65% lifetime risk of developing breast cancer and for women with a BRCA2

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 8
August 2017

mutation, the risk is 45%. The two genes mutations often tends to develop in younger women. The two major risk factors are gender (being a woman) and age (growing older) [3].

The data analysis plays a vital role to find the major causes of the disease. It has a great deal of impact in the area of Breast cancer. At present people getting affected with breast cancer is increasing due to various factors. Everyday the voluminous amount of data being generated from multiple departments of the hospital. The generated data must be converted into useful information and knowledge for the use of society at large and it can be used for further research. Normally in hospitals data can be stored in multiple databases[4]. A repository of multiple heterogeneous data sources from different departments in the hospital such as mammography division, scanning division and laboratory organized under a unified schema at a single site in order to facilitate chief doctor decision making to treat and monitor the patient health, controlling the level of disease by screening, chemotherapy, radiation therapy, hormone therapy, mastectomy and operational procedures. These measures are leading to complete eradication of cancer or controlling by increasing the patient survival time [5].

II. FEATURE SELECTION AND CLASSIFICATION

The breast cancer prediction (Kharya, S., & Soni, S. 2016) was formulated based on Naïve bayes classifier. Here, Naïve Bayes was incremented to analyze in a improved manner for getting the classification performance of Naïve Bayes by adding weighted concept with the Naïve Bayes. Here, the Naïve Bayes was incremented by assigning weight to attributes of available breast cancer data in the database. But in this approach the primary data which can give only limited details about the patients. The patient details are getting utmost importance for the prediction of the breast cancer. At this juncture feature selection was not performed because it may lead into high dimensionality problem. Here, the different heterogeneous data sources are collected from different departments in the hospital which can provide the entire details of the patients in which some data are directly collected from patients. The breast cancer risk of the patient is critically analyzed and calculated by applying data through medical analysis, BIRAD Score, Medical History, Social History and Family History to the system. The output of the system which suggests the doctor for giving appropriate medical advice to the patient. However, it needs advanced approach like feature selection and classification method for the reduction of high dimensionality problem. In this research, feature selection and classification methodologies are propose to predict the breast cancer in the patient body. Here, the discriminative features from the available dataset are extracted by applying firefly optimization algorithm. The accurate combinations of closely interrelated features are obtained by the firefly optimization algorithm. The breast cancer recurrence prediction is enhanced by firefly feature selection algorithm. The identified features are used for training classifier to analyze and predict the test datasets. The ensemble learning classifier Bagged Decision Tree (BDT) is used for classification. Bagging is generally applied to decision tree methods, anyway it can be used with other classifiers techniques. Finally bagged tree, resample the training dataset into multiple times, and builds a decision tree model from each data set and then amalgamate these models together for a resultant classifier.

III. FIREFLY ALGORITHM

The main purpose for using firefly's flash is to act as a signal system to attract other fireflies. This algorithm formulated by Xin-She Yang by articulating the following factors:

1. Each and everyone fireflies are unisexual, so that any one firefly can be attracted to all other fireflies;
2. Attractiveness is proportional to their brightness, the less bright having attraction towards the brighter one, but the intensity decreases when their mutual distance increases;
3. If at all there is no brighter than a given firefly, it will pass randomly.

The brightness related to the objective function. The maximization problem states that the brightness is directly proportional to the objective function’s value. It can also defined based on the

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 8
August 2017

fitness function which is used in the genetic algorithms. There are two important factors in the Firefly algorithm, the first one is light intensity variation and the second one is attractiveness among the flies. The objective function derived based on assumed that attractiveness of firefly by its brightness. When the fly in the particular location x and the available brightness I of a firefly can be derived as I(x) f(x) for a maximization problem.

The attractiveness β is judged by other fireflies and it will differ with the distance (i,j) between two fireflies (firefly i and firefly j). As per the formulated law, the light intensity decreases with distance from its source and light is also absorbed by air. The attractiveness should be allowed to vary with varying degree of absorption. In the normal method, light intensity I(r) increases according to inverse square law.

Firefly algorithm()

Objective function f(x),x=(x1,x2,…,xd)

Initialize the number of fire flies xi(i= 1,2,. ., n)

Formulate light absorption coefficient Û

While (t<Max.Generations)

For i=1 : n(all n fireflies)

For j=1: n

Intensity of the Light I i at xi is determined by f(xi)

If(Ii >Ij)

Move firefly i towards j in all dimensions

Else

Move firefly i randomly

End If

Attractiveness changes with distance r via exp[-Ûr2]

Determine new solutions and change light intensity

End for j

End for i

Rank the fireflies based on light intensity and find the present best pair

End while

IV. IMAGE BASED OPTIMIZATION TECHNIQUE

The Firefly and Bagged Decision Tree (BDT) methods can analyze and predict based on the data collected from multiple sources. The image is the crucial factor for predicting breast cancer and prediction based on features derived from breast images. The tumor patch is found by the segmentation phase by implementing morphological closing operation and image gradient technique to find the Region Of Interest (ROI). The Max-Mean and Least-Variance technique can be applied to create a rectangular window around the output region to find the malignant tissues. The calculated values of Max-Mean and Least-Variance are derived as features for prediction of breast cancer. The averaging filter and threshold operation is used on actual original input image which gives the output in the malignant cancer region area. In the process of feature selection the same firefly algorithm automatically selects the threshold parameter which is the main deciding factor. If there are multiple independent breast images [7], the parameter adaption using firefly algorithm improves the segmentation process. In general the Bagged Decision Tree (BDT) is used for purpose of classification. The improvement analysis for predicting breast cancer based on optimization approach. The breast cancer prediction based on accuracy and it can be evolved by adding the features of breast cancer data and image. By processing such heterogeneous data, the needed features are identified and selected and then combine by using firefly optimization algorithm. The selected features from both data and image considered as inputs to the Bagged Decision Tree classification for predicting the breast cancer recurrence efficiently and effectively. So the predication methodology can be improved because of using heterogeneous dataset. The breast cancer is a complex disease because of multiple parameters need to be analyzed, the improvement still required to get utmost precision to predict the breast cancer. An enhanced classification is proposed by adding statistical features along with features from breast cancer data and breast images. Here the data selection based on major and minor parameters which causes breast cancer. In the statistical analysis the features such as skew, kurtosis and etc

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 8
August 2017

are manipulated and fused with available features. These statistical parameters are supported to improve classification accuracy level. By using firefly optimization algorithm, most of the discriminative features were selected to improve the accuracy. The selected features again passed as input to the Bagged Decision Tree for classification purpose to predict the breast cancer recurrence. Hence, our proposed work improves the accuracy of selection and classification for predicting the breast cancer recurrence than other known methods.

V. DATASET AND ACCURACY PARAMETER

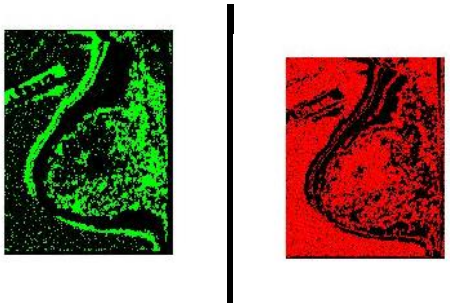
The data has been collected from one of the leading hospital in Bangalore. The breast cancer data consists of 30 attributes and 1 class attribute in which some of the attributes are primary and remaining are secondary [8]. There are multiple details are recorded such as LVI-nodes, EIC, tumor type, tumor size [6], surgery happened or not, Chemotherapy, Radiotherapy, Hormone Therapy, Follow-up, Last to follow up and alive or dead at present. There are 300 images breast cancer images are used in this analysis.

Accuracy or classification accuracy is termed as the ability of the classifier and is calculated by,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

VI. RESULTS

The experiment has been performed by using Matlab. Matalb tool is independently used for data and image processing. In the following images the input cluster images processed for multiple levels like noise removal, classification and segmentation. The input is data and image and the output is the processed data as well as image. But the comparison is required for multiple algorithms to bring out result with high level accuracy.



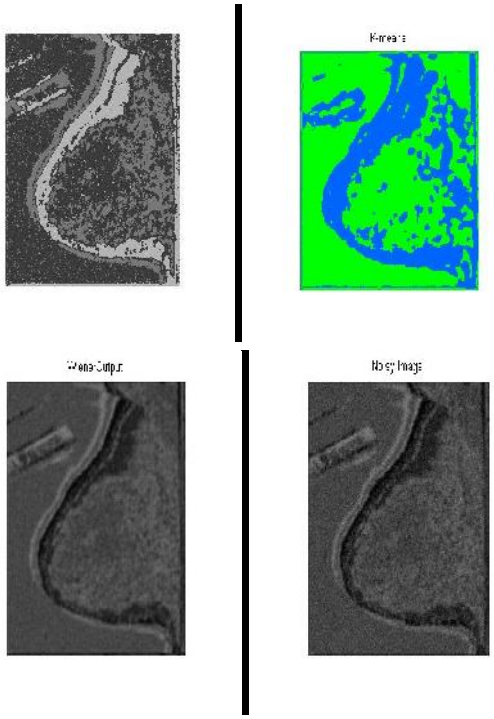


Fig1. Breast cancer images

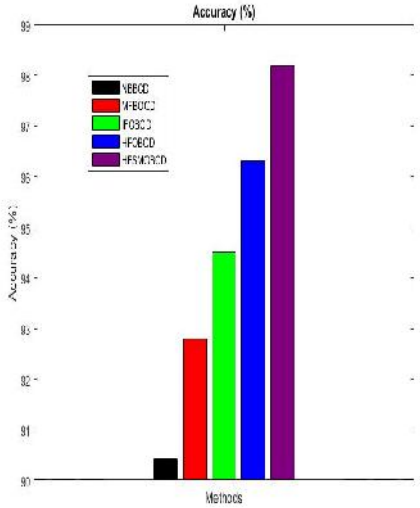


Fig 2.Comparison of Accuracy for Breast Cancer Detection

VII. CONCLUSION

The experiment has been performed by using Matlab. Matalb tool is independently used for data and image processing. In the fig.1 there are the multiple images which was applied for the processing level. The fig.2 shows that comparative result by using five algorithms based on the

accuracy parameter. The five algorithms are used here, Naïve Bayes based Breast Cancer Detection (NBBCD) [10], Multiple factor (primary and secondary) based optimized Breast cancer detection (MFBOCD), Image feature based optimized Breast cancer detection (IFOBCD), heterogeneous feature based optimized breast cancer detection (HFOBCD) and heterogeneous feature with statistical measure based breast cancer detection (HFSMOBCD) approach. In this graph, x-axis as multiple classification methods such as NBBCD, MFBOCD, IFOBCD, HFOBCD and HFSMOBCD and y-axis as accuracy rate. The accuracy differs from one method to other method because of the complexity of the breast cancer disease. The accuracy comparison graph gives a crystal clear idea, the HFSMOBCD method has more efficient than others because of maximum accuracy. The main reason for doing this research to find out the efficacy of the tumor cells like benign or malignant where as finding the seed cell and relative cell based on firefly algorithm is important task because of the disease complex in nature. The performance of any algorithm rely upon the parameters used for the process. So each process level the accuracy can be strengthened. The prediction of the breast cancer using enhanced different techniques changes from time to time because of technological advancement. Initially, the classification method is enhanced by focusing on the heterogeneous data, feature selection based on firefly optimization algorithm and classification based on Bagged Decision Tree (BDT) [9]. In these methods the firefly optimization algorithm identifies the more reliable features from breast cancer data and image. Ultimately these features are used in BDT to classify the breast cancer data, in order to predict the breast cancer from breast images.

REFERENCES

[1] Hanan Alshanbari, Saad Amain, James Shuttelworth, Kaled Slman and Shadi Muslam , Automatic Segmentation in Breast Cancer Using Watershed Algorithm, International Journal of Biomedical Engineering and Science (IJBES),Vol.2, No.2, April 2015

[2] http://www.breastcancerindia.net/statistics/trends.html

[3] http://www.breastcancer.org/symptoms/understand_bc/statistics.

[4] Zaheeruddin, Z.A. Jaffery and Laxman Singh, Detection and shape feature extraction of Breast tumor in mammograms, Proceedings of the World Congress on Engineering , Vol II, WCE 2012, London, U.K.

[5] 5. S. Saheb Basha, K.Satya Prasad, Automatic Detection of Breast Cancer Mass in Mammograms using Morphological Operators and Fuzzy C-means clustering, Journal of Theoretical and Applied Information Technology.

[6] Rahmatinia, S., & Fahimi, B. (2017). Magneto-Thermal Modeling of Biological Tissues: A Step Toward Breast Cancer Detection. IEEE Transactions on Magnetics, 53(6), 1-4.

[7] Sabeena, K., Nair, M. S., & Bindu, G. R. (2017). A Multi-Classifier System for Automatic Mitosis Detection in Breast Histopathology Images Using Deep Belief Networks. IEEE Journal of Translational Engineering in Health and Medicine.

[8] Beura, S., Majhi, B., Dash, R., & Roy, S. (2015). Classification of mammogram using two-dimensional discrete orthonormal S-transform for breast cancer detection. Healthcare technology letters, 2(2), 46-51.

[9] Johra, F. T., & Shuvo, M. M. H. (2016, September). Detection of breast cancer from histopathology image and classifying benign and malignant state using fuzzy logic. In Electrical Engineering and Information Communication Technology (ICEEICT), 2016 3rd International Conference on (pp. 1-5). IEEE.

[10] Kharya, S., & Soni, S. (2016). Weighted naive bayes classifier: A predictive model for breast cancer detection. International Journal of Computer Applications, 133(9), 32-7.

Mr.Saravanakumar has been working as a associate professor at Christ University, Bangalore, India. He has completed MCA,M.Phil,MBA,M.Tech and having teaching experience of nearly 18 years. His area of interests are Data analysis, Medical Image Processing , Analysis and Design of Algorithms, Distributed Database and Operating System.

Dr Arthanari holds a Ph.D in Mathematics from Madras University as well as Masters degree in Computer Science from BITS, Pilani. He holds patent issued by the Govt. of India for the invention in the field of Computer Science. He has directed teams of Ph.D researchers and industry experts for developing patentable products. He teaches strategy, project management, creative problem solving, innovation and integrated new product development for last 35 years.