Proactive Prediction of Breast Cancer by Implementing Crow Search Optimization

Dr. B. Gomathi
Assistant Professor,
Department of Computer Science,
Shri Shankarlal Sundarbai Shasun Jain College for Women,
Chennai, Tamilnadu, India.
Email: gomathiphd2k15@gmail.com

1. Abstract

Cancer is one of the deadly diseases with high rate of lossof human life. Breast Cancer is becoming very common in Women and as per statistics it is curable when detected at early stages. One of the tools to identify is classification algorithm. Several procedures are available but the key issue is to identify a method that can provide high accuracy and greater precision. In order to proactively identify, in this paper a comparison of different classification algorithm is attempted. Classifiers like Naive Bayes, Best First Search, Random Search and Genetic search are compared and our novel approach of integrating Naive Byes with Best First search and Crow optimization provided better results. This paper focus on fusion of optimization with classification to enable early prediction ,and data set is used from UCI repository.

Keywords: Crow search algorithm, classification, Feature selection, Naive Bayes, Best first search, Random search and Genetics search.

The disease that causes lot of human loss especially in Women is Breast cancer. There is always a very good chance for long term survival of Breast cancer patient if diagnosis is done at early stage. The cause for the disease is uncertain but scientist confirm that riskfactors might be due to ageing, obesity, hereditary, menopause, consumption of alcohol etc [1]. Predictive algorithm plays a vital role in all industries like ERP, Healthcare,insurance, Airlines, Media and also in security exchanges [2]. Health care industry provides lot of data related to patients, details pertaining to conclusion of disease and also steps to predict and prevent them. For characterization an index is provided and this contains different attribute including both numerical and category data. Attributes like size of the tumour, age, Deg-Malig, Menopause belongs to numerical type and properties like Irradiant, Breast, Node-cape, Breast-Quad, Class belong to categorical type.

Classification plays a vital role in excruciating a dataset into mutually select groups, called a class, based on suitable attributes. This disease affects different age groups with varied symptoms for example if it is due to milk ducts then it is named as ductal carcinoma, and if cells are found in lobules it is referred as lobular carcinoma.This paper illustrates about the classification techniques and results prove that fusion of Naive Bayes with BFS and Crow search optimization provides better results.

3.Review of Literature

Hybridapproach proved better with respect to end stage renal disease and this was illustrated by the author You-Shyang Chen [4]. One of the AI search strategyis Best First Search andit

allows back tracking along the search path. The entire search space will be utilised for the time given and it is always essential to give stop criterion. [5][6].
The author Shashikant Ghumbre et al[7]., provided a survey of different methods for diagnosis and proved that deep learning gives better result when compared with genetic searches. Comparison of different ML algorithms using Wisconsin Data was done by the author Hiba Asri et al [7] and he proved SVM provides better results.Optimization yields better results and comparison of few techniques like Whale Optimization Algorithm(WOA), Dragonfly Algorithm (DA) and Moth-Flame Optimization (MFO) algorithm was done by the author Ayat Ali Saleh[8].

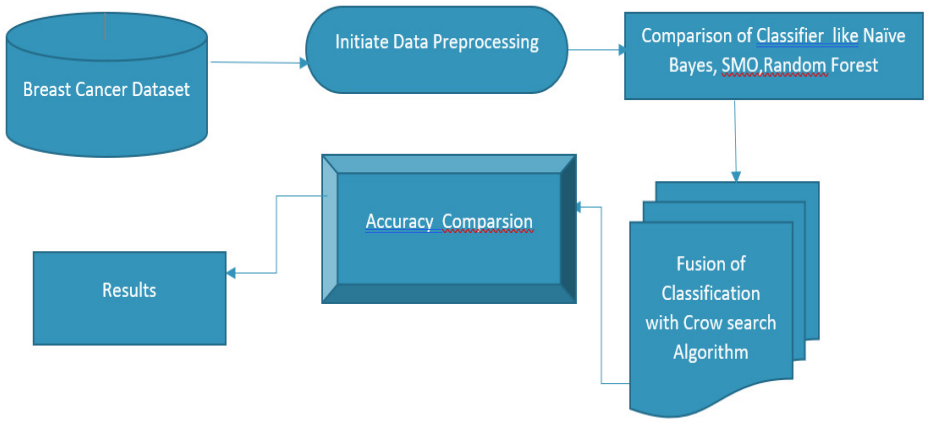
4.Proposed Methodology

In our approach performance of algorithms namely BFS, Naive Bayes, Genetic search and Random search is implemented using medical data from UCI repository. The attributes used are derived from medical reviews with ten features and are given in the following table 1.

Table 1: UCI Dataset of Breast Cancer

Attributes Name	Attribute Type	Description
Age	Numeric	Age (years)
Inv-Nodes	Numeric	0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39
Node-Caps	Discrete	yes, no.
Menopause	Numeric	lt40, ge40, premeno
Deg-Malig	Numeric	1, 2, 3.
Tumor-Size	Numeric	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59
Breast	Discrete	left, right
Breast-Quad	Discrete	left-up, left-low, right-up, right-low, central.
Irradiat	Discrete	yes, no.
Class	Discrete	no-recurrence-events, recurrence-events

The basic idea is to find Precision, Recall and FMeasure of the various methods and compare them to find the best solution. Then the results are combined with Crow search algorithm and the fusion approach provided optimal results. The proposed system architecture is mentioned in the below figure 1.



.

Figure 1: Proposed System Architecture

4.1 Fusion Algorithm of Naive Bayes with BFS and Crow Search

Step 1: Initialise the dataset .

Step 2: Set Start and Stop State.

Step 3: Assign BEST state to Start.

Step 4: Proceed with the subset till BEST position is reached.

Step 5: close the list and move with next element.

Step 6: With the new data check for both training as well as test data.

Step 7: Based on the conditions arrive at probablities.

Step 8: Assign the subset with respect to crow position.

Step 9: Find the new position and update with the current state.

Step 10: Check for the best state otherwise go to step 2

Accuracy is based on true positive and negative values for a given set of data. The below table 2 show accuracy of the various algorithm and the Figure 2 represents the same.

Table 2:Comparison of Accuracy

Method	Accuracy (%)
NB	74
GNS -NB	75
RNS- NB	78
BFS - NB	84
BFS-NB-Crow Search	87

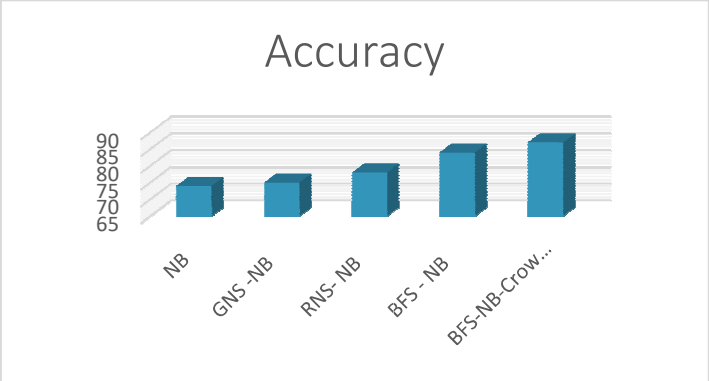


Figure 2: Comparison of Accuracy

The precision and recall value of our fusion method of integration BFS-Naive Bayes with Crow search optimization yields better result when compared to other methods. This is depicted in the below table 3.

Table 3: Comparison of Precision, Recall and FMeasure

Method	Precision	Recall	FMeasure
NB	**86**	**71**	**70**
GNS -NB	**84**	**69**	**67**
RNS- NB	**74**	**64**	**62**
BFS - NB	**83**	**79**	**80**
BFS-NB-Crow Search	**90**	**82**	**74**

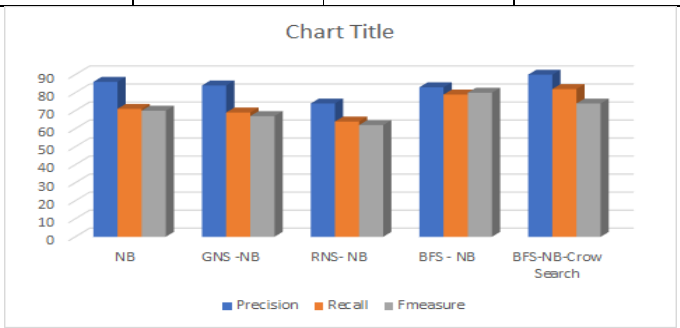


Figure 3: Comparison of Precision, Recall and Fmeasure

5.Conclusion

In this paper performance was compared with medical dataset in order to find the efficient method enabling proactive deduction of breast cancer. The result show that fusion of Best first Search with Naive Bayes and crow search optimization yields better result. In future we can also implement fuzzy rules along with other optimization algorithm for better results.

6.References

1. Saheb Basha, Dr.K.Satya Prasad ,"Automatic Detection Of Breast Cancer Mass In Mammograms Using Morphological Operators And Fuzzy C-Means Clustering".Journal of Theoretical & Applied Information Technology
2. Sitar-Taut, V.A., et al, Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.

3. You-Shyang Chen, Modelling hybrid rough set-based classification procedures to identify haemodialysis adequacy for end-stage renal disease patients, Computers in Biology and Medicine, 2013, vol. 43, pp. 1590–1605.

4. Hall, Mark A. and Lloyd A. Smith., "Feature subset selection: a correlation-based filter approach", 1997.
5. Hall. M, "Correlation based feature selection for machine learning", Doctoral dissertation, University of Waikato, Dept. of Computer Science, 1999.
6. Pallabi Borah et al., "A statistical feature selection technique", Netw Model Anal Health Inform Bioinforma, pp-3-55, 2014.
7. Hiba Asri,HajarMousannif,Hassan Al MoatassimeT,homasNeol," Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis", Volume 83, 2016, Pages 1064-1069
8. Ayat Ali Saleh ; Al-Attar Ali Mohamed ; Ashraf Mohamed Hemeida ; Abdalla Ahmed Ibrahim ,"Comparison of different optimization techniques for optimal allocation of multiple distribution generation",2018, International Conference on Innovative Trends in Computer Engineering (ITCE)
9. Ahmed M. Anter, Aboul Ella Hassenian, Diego Oliva ,"An improved fast fuzzy c-means using crow search optimization algorithm for crop identification in agricultural", Expert Systems with Applications · October 2018.