PROGNOSTICATION NOVEL FEATURE ENGINEERING FOR MALIGNANT BREAST CARCINOMA DIAGNOSIS

Dr.S.SARANYA
Assistant Professor, Department of Computer Science with Cognitive System and Artificial Intelligence & Machine Learning,
Hindusthan College of Arts & Science, Coimbatore, Tamil Nadu, India.

Dr.P.DEEPIKA
Associate professor, Department of Artificial Intelligence and Machine Learning,
Dr. N.G.P Arts & Science College,. Coimbatore, Tamil Nadu, India.

ABSTRACT

Breast carcinoma is the process of growth abnormal lump in the breast. It begins in the cell and it scatter all over the body. Breast carcinoma is first detected by diagnostic tests and procedures that may be used to confirm the presence of cancer and to determine if it has been Spread or not. In Worldwide, breast carcinoma is the majority frequently recognized mortal cancer in women and the leading source of carcinoma death among the women. Early analysis of carcinoma on discovers symptomatic patients as premature as feasible so the medical supervisor has able to provide the successful treatment. When the carcinoma has diagnosis on later then the possibility of survivance is less, cost of the treatment will be high. This research has focused on the type of the carcinoma at the early stage of cancer to save the life of the patients. Breast carcinoma types of benign/malignant prediction throughout the proper machine learning techniques. WBCO dataset are utilized in the study. The breast carcinoma mortality ratio is increasing has to be noticeable. This research actuates the deep learning techniques in the breast carcinoma prediction. The feature selection is based on the proposed OPFF algorithm to select the features of high significance in the algorithm which helps to improve the classification accuracy. The classification algorithm of XGBMLP are based on the deep learning neural networks to improve the classification accuracy with and without feature selection and concluded the proposed classification algorithm have higher accuracy when compared to the various existing algorithm.

Keywords: Breast carcinoma, Prediction, Feature Selection, Deep Learning.

INTRODUCTION

In the year of 2020, the 2.5 millions women's were analyze the Carcinoma disease in the world and among 685 000 life expired. In the year of 2021, 7.9 millions women's were analyzed the breast Carcinoma disease due to low awareness of carcinoma treatment at early stage. So the world most common disease is the breast carcinoma. The women's after the age of puberty the carcinoma has been affected mostly and many women's were expired. Carcinoma cells begins in the breast and it gradual pass the Carcinoma cell all over the body. A usual cell grows in the body but the unwanted cells with previous cell formation will forms as Carcinoma cells. The carcinoma has the couple of types. The first type is malignant and the second type is benign. The second type of the cancer will not

separate all over the body but the malignant is the life expire disease.

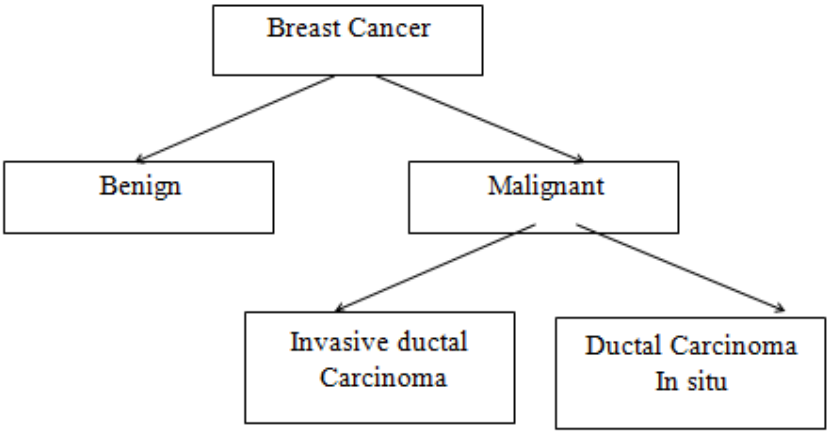

Figure 1 Taxonomy of Breast Carcinoma

Breast Carcinoma is the 5th common cancer for the death rate in the world the breast cancer types of benign and malignant and the malignant classify in to couple of types they are invasive ductal carcinoma and ductal carcinoma. This Invasive ductal carcinoma is the most life treating disease. Breast cancer can be classified as follows:

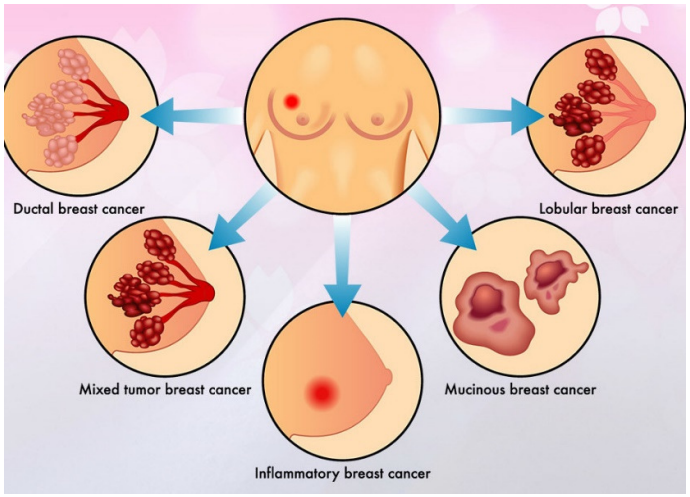

Figure 2: Exemplification of the breast cancer

DATASET DESCRIPTION

Breast cancer benign/malignant prediction is enacted and evaluate through the suitable machine learning techniques in this research. Wisconsin original breast cancer dataset: This Wisconsin original breast cancer dataset.

This dataset contains the attributes including the id, clump thickness ranging from 1 to 10, uniformity of cell size and cell shape from 1 to 10, marginal adhesion, epithelial size of single cell, bare nuclei, bland chromatin, normal nuclei, mitoses ranging from 1 to 10, class variable indicating the benign and malignant breast cancer, value 2 representing benign and 4 representing malignant. This dataset has 699 instances of breast carcinoma patients and 11 attributes. It has 16 missing data values, these are deleted manual so now totally 683 records. In this 699 instances 456 are belongs to malignant cancer and 241 are belong to benign cancer.

PROPOSED FEATURE SELECTION

In this research the feature selection has been initiated to improve the accuracy of classification and also it will concentrate on high significance features. It helps to reduce the dimensionality of dataset. It avoids the time consuming for the irrelevant features. This research has proposed the optimal propagation fireflies – OPFF for the feature selection.

Firefly algorithm is a combinatorial optimization algorithm. The feature selection based on the firefly's algorithm has been usually accepted strategy. in the existing firefly is based on the brightness of attractive other firefly and hence the convergency process time may be slow for the problem. To overcome this, the research proposed a novel firefly algorithm based on the optimal propagation scenario. Firefly present in the population as x=1,2,3…. N this firefly has been attractive by the other fireflies based on the brightness. In enhanced optimal propagation scenario, the limit is diminishing through an optimal propagation fitness function. This fitness function will get the best firefly along with the optimal firefly. The following is the formula for computing the optimal propagation fitness function.

$$OPro(f(x)) = x_b + opt\,(x_i + x_j) \text{ and } opt(x) = \frac{|x_b|}{|x_b + opt\,(x_i + x_j)|}$$

The opt is the function to compute the x_i and x_j are the optimal estimate for the fireflies. So once the estimate value has propagated then for present iteration that corresponding firefly will not be taken again. . The x_b is the best firefly which will be hold for the complete iteration. . But for the optimal fireflies is different. Once they are propagated they will be proceed to the permanent places in the dataset and also it gets ranked. For the next iteration the optima flies will be selected based on the fitness function. Again now the present fireflies under consideration are propagated they will be proceed to permanent places in the dataset and get ranked. So the best firefly selection process is the one time processing whereas the optimal firefly selection is iterative progression among the ranking. This research makes us of the optimal firefly algorithm for the feature selection process.

The attributes selected after feature selection is 9 features out of 11.. The search space illustration is below figure.

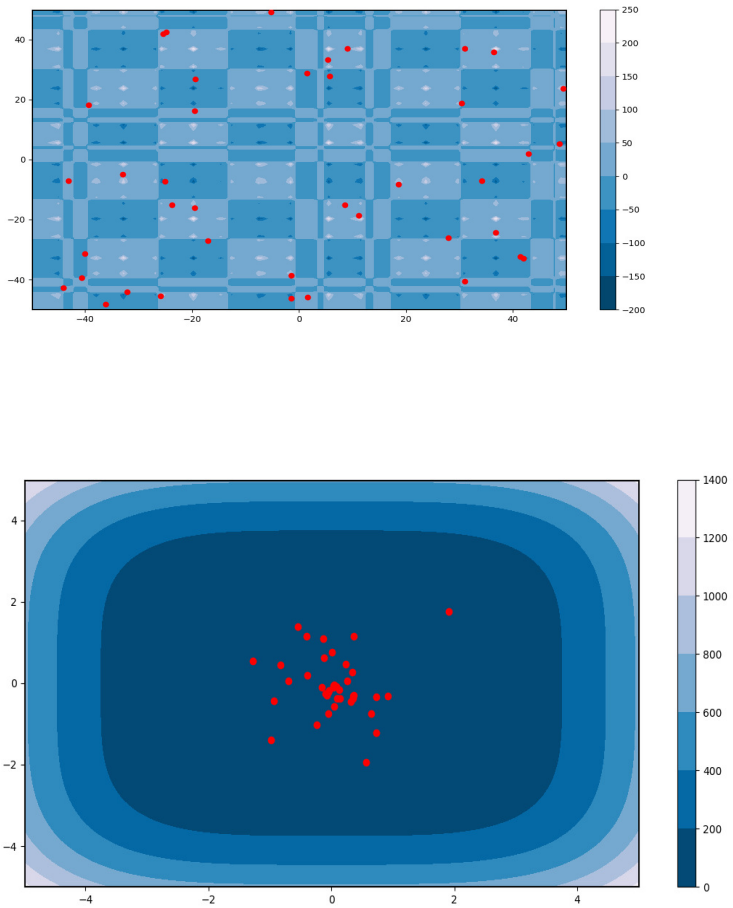

Figure 3: Visualizing the features correlation selected as the fireflies propagated

PROPOSED XGBMLP ALGOROTHIM

The Proposed method of XGBMLP Places the optimization function extreme gradeint boosting to get the deduction of features from the XGB trees as been deployed. It will makes the update of every layer of neural network throuth the way of updation thoiugh XGB trees, the feature significance revelance will be the results. In the health care data analysis the accuracy acting as a main task, so the life of the pateients has been saved by analysing the disease at early stage. Multilayer perceptron (MLP) is a feedforward artificial rtificial networks and deep learning method which produces the output from set of the inputs. Extreme gradient boosting will achieve well in boosting and also discard the overfitting with significant inference retrieval. Multilayer perceptron (MLP) is a feedforward artificial rtificial networks and deep learning method which generates set of outputs from the set of set of inputs. Extreme gradient boosting perfroms well in boosting and it also avoids the overfitting with significant inference retrieval. Feature interpretation were used through the neural network.

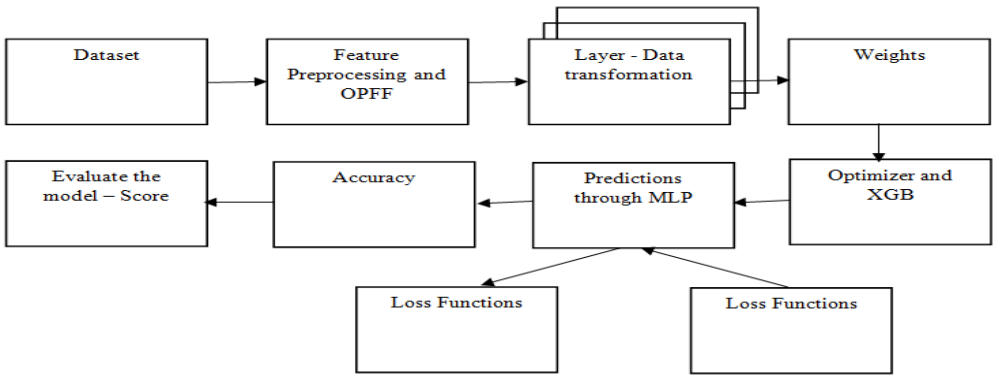

Figure 4: Workflow of EGBMLP

MLP will be well suitable for non-linear seperable problems. In MLP the hidden layer of the network the neurons has not been directly connected to output layer. The hidden layer will be acted as a separate hyperplanes. This can improve the detachment capacity of the network. Hhence here the hidden layer hyperplane the extreme gradient boosting can be utilizing as the final layer for analyzing the class. Activation function will helps to the artificial neural network to recognize and get deduction from complex patterns. This study has instigates the various activation functions including: Relu, Logistic and tanh. The class prediction layer in the neural network is proceed using the Extreme Gradient Boosting which possesss inextricable part in predicting the model perfromance.

RESULTS & DISCUSSION

Feature selection acting a very important role in the improvement of the results in the accuracy for the WBCO datasets. Machine and deep learning methods comparison and summary, brief of the algorithms established in the study for comparing with the proposed algorithms. It also extend the deep learning based method named, Extreme Gradient Boosting based Multi-Layer Perceptron, XGBMLP in comprehensive manner with proper exemplification. Experimental settings are tabulated and results are extend in neat manner. The results are compared with different existing algorithms to expose the performance of the proposed algorithm.

Table 1: Overall performance comparison of the proposed system, XGBMLP with other MLT for WBCO Dataset

Method	**Precision**	**Recall**	**F-Score**	**Accuracy**	**Method**	**Precision**	**Recall**	**F-Score**	**Accuracy**
WBCO Dataset – Feature selection					**WBCO Dataset - No Feature selection**				
SVM	0.96	0.97	0.96	97.81	SVM	0.86	0.87	0.86	86.91
KNN	0.97	0.96	096	96.35	KNN	0.81	0.81	0.80	85.32
LR	0.97	0.98	0.98	97.81	LR	0.95	0.95	0.95	94.21
NB	0.94	0.96	0.95	94.89	NB	0.91	0.90	0.91	91.25
MLP	0.92	0.93	0.92	92.70	MLP	0.93	0.93	0.94	93.27
XGBMLP	**0.95**	**0.95**	**0.98**	**98.70**	**XGBMLP**	**0.95**	**0.95**	**0.95**	**97.7**

It is comparatively notice that classifiers have been elevated accuracy after features selection among all the existing methods. Feature selection gives an important impact in improving the accuracy of the classifiers for breast carcinoma detection using the WBCO. Breast cancer classification using WBCO with features selection is enhancing the accuracy. XGBMLP performs well when compared with the proposed method with feature selection 98.70% higher accuracy than the other algorithm and without feature selection 97.7% accuracy.

CONCLUSION

Deep learning produces the hopeful aspects in the prediction of the breast cancer disease. This study instigates the deep learning techniques in breast cancer prediction. Feature selection acting a very important role in the improvement of the results in the accuracy for the WBCO datasets. . The accuracy with and without feature selection shows the less difference. But it is witnessed that the proposed method XGBMLP gives overall higher accuracy 98.70% for the classification of the breast cancer prediction using WBCO dataset.

REFERENCES

1. Breast Cancer Wisconsin (Diagnostic), Database, accessed at : https://datasetsearch.research.google.com/search?query=breast%20cancer&docid=L2cvMTFqOWI0bjl4ZA%3D%3D, updated on dec 8 2021
2. D. Bazazeh and R. Shubair, "Comparative study of machine learning algorithms for breast cancer detection and diagnosis," 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), 2016, pp. 1-4, doi: 10.1109/ICEDSA.2016.7818560.
3. Ms P Deepika, Ms S Sasikala, S Jansi, Ms A Kiruthika, Ms S Saranya(2016). Accelerating Unique Strategy for Centroid Priming in K-Means Clustering. IJIRST –International Journal for Innovative Research in Science & Technology| Volume 3
4. P Deepika, S Saranya, Dr S Sasikala, Haritha RA, Aravind G Nair(2018). The Study of Fuzzy Cognitive Maps in Identification and Prediction of Disease. IOSR Journal of Engineering.

5. A.Kiruthika P.Deepika, S.Saranya, Dr.S.Sasikala(2018). A Noval Clasiification and Prediction Algorithm For Heart Disease Identification. International Journal of Engineering Science Invention (IJESI), vol 7, Pages: 12-17
6. Z. Nematzadeh, R. Ibrahim and A. Selamat, "Comparative studies on breast cancer classifications with k-fold cross validations using machine learning techniques," *2015 10th Asian Control Conference (ASCC)*, 2015, pp. 1-6, doi: 10.1109/ASCC.2015.7244654.
7. Dr.S.Sasikala S.Saranya , P.Deepika(2018). Comprehensive Review on Heart Disease Prediction Using Optimization Techniques. EPRA International Journal of Research and Development (IJRD) Peer Reviewed Journal
8. S Saranya, P Deepika, S Sasikala(2018). Comprehensive Review on Heart Disease Prediction using Optimization Techniques. EPRA International Journal of Research and Development (IJRD), vol 5
9. S Sasikala, S Jansi, S Saranya, P Deepika, A Kiruthika(2017). Anticipating the chronic kidney disorder (CKD) using performance optimization in adaboost and multilayer perceptron Imp
10. E. Yavuz, C. Eyupoglu, U. Sanver and R. Yazici, "An ensemble of neural networks for breast cancer diagnosis," 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 538-543, doi: 10.1109/UBMK.2017.8093456.