BREAST CANCER DETECTION USING MACHINE LEARNING-THE SURVEY

Prof.Mrunalinee Patole

Rutuja Deshmukh[1],Shrawani Harkal[2],Rohit Deshmukh[3],Rushikesh Chaudhari[4]

[1]Department of Computer Science, RMDSSOE, Pune,India

[2]Department of Computer Science, RMDSSOE, Pune,India

[3]Department of Computer Science, RMDSSOE, Pune,India

[4]Department of Computer Science, RMDSSOE, Pune,India

ABSTRACT :

Breast cancer is the leading cause of death among women. Several types of research have been done on early detection of breast cancer to start treatment and increase the chance of survival. Most of the studies concentrated on mammogram images. However, mammogram images sometimes have a risk of false detection that may endanger the patient's health. It is vital to find alternative methods which are easier to implement and work with different data sets, cheaper and safer, that can produce a more reliable prediction. This paper proposes a hybrid model combined of several Machine Learning (ML) algorithms including Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Decision Tree (DT) for effective breast cancer detection. This study also discusses the datasets used for breast cancer detection and diagnosis. The proposed model can be used with different data types such as image, blood, etc.

Keywords *–Accurat ,Benign,Malignant,Breast Cancer, DNN, Earlierdetection, GLCM, Mamograms, Wavelettransform*

1. Introduction

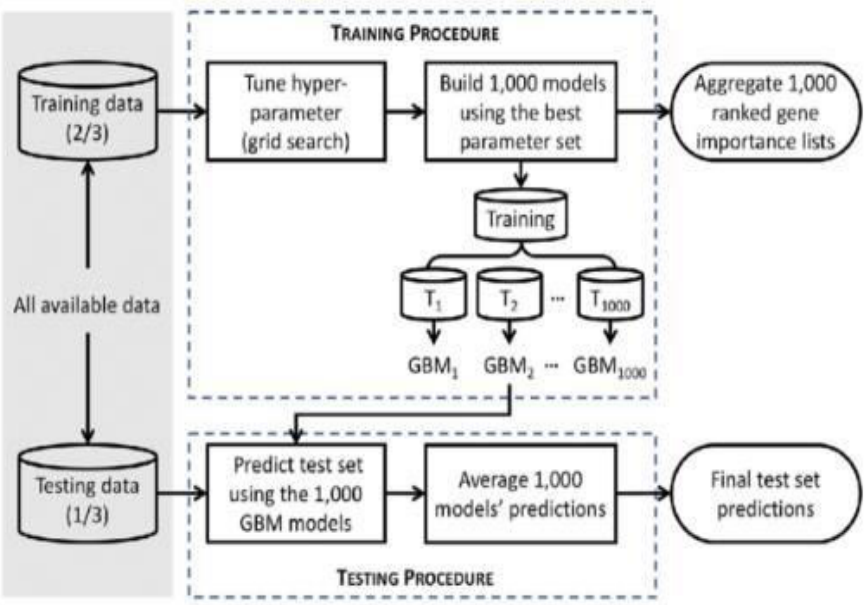
The major health issue that arises these days had led to much advancement in the medical field, yet there are certain diseases which remains incurable. There are so many cancers, among which Breast cancer is the most important cause of cancer death among women. Early detection of breast cancer aids for early diagnosis and treatment, because the prognosis is very important for long term survival. Machine Learning (ML), is a subfield of Artificial Intelligence (AI) that allows machines to learn without explicit programming by exposing them to sets of data allowing them to learn a specific task through experience. Over the last few decades, ML methods have been widespread in the development of predictive models in order to support effective decision-making. In cancer research, these techniques could be used to identify different patterns in a data set and consequently predict whether a cancer is malignant or benign. The performance of such techniques can be evaluated based on the accuracy of the classification, recall, precision, and the area under the ROC. By studying these papers we created a machine learning model which detects whether the cancer is Malignant or Benign.

2. Machine Learning Methods

Machine Learning is a process that machines (computers) are trained with data to make the decision for similar cases . ML is employed in various applications, such as object recognition, network, security, and healthcare. There are two ML types i.e. single and hybrid methods like ANN, SVM, Gaussian Mixture Model (GMM), K-Nearest Neighbor (KNN), Linear Regressive Classification (LRC), Weighted Hierarchical Adaptive Voting Ensemble (WHAVE), etc. Following are the used ML algorithms:

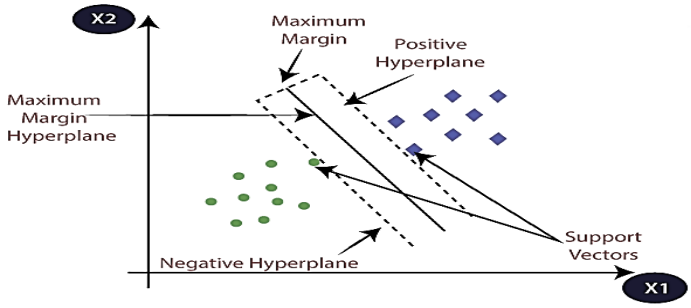
2.1 XGBoost Classifier :

eXtreme Gradient Boosting or XGBoost is a library of gradient boosting algorithms optimized for modern data science problems and tools. Some of the major benefits of XGBoost are that it's highly scalable/parallelizable, quick to execute, and typically outperforms other algorithms and used a more regularized model formalization, to control over-fitting, which gives it better performance.



2.2 Support Vector Machine (SVM)

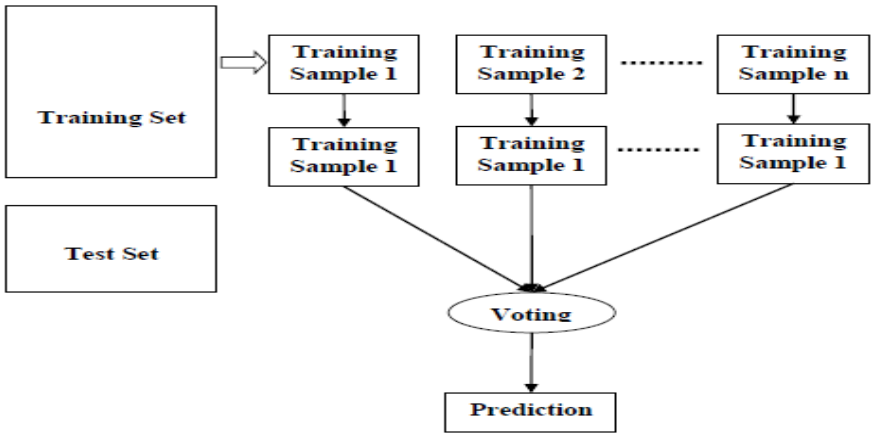
SVM is a supervised pattern classification model which is used as a training algorithm for learning classification and regression rule from gathered data . The purpose of this method is to separate data until a hyperplane with high minimum distance is found. SVM is used to classify two or more data types. SVM include single or hybrid models such as Standard SVM (St-SVM), Proximal Support Vector Machine (PSVM), Newton Support Vector Machine (NSVM), Lagrangian Support Vector Machines (LSVM), Linear Programming Support Vector Machines (LPSVM), and Smooth Support Vector Machine
(SSVM).



2.3 Random Forest (RF)

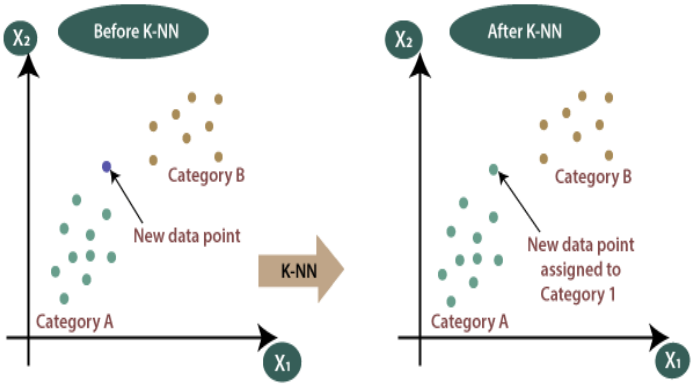
Random Forest classifier is a learning method that operates by constructing multiple decision trees and the final decision is made based on the majority of the trees and is chosen by the random forest. It is a tree-shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, instance, or reaction.

Using of Random Forest Algorithm is one of the main advantages is that it reduces the risk of over fitting and the required training time. Additionally, it also offers a high level of accuracy.



2.4 K-Nearest Neighbour

It is one of the simplest Machine Learning algorithms based on Supervised Learning technique. And assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. It stores all the available data and classifies a new data point based on the similarity and easily classified into a well suite category by using K- NN algorithm.



Comparative Study of Machine Learning Algorithms for Breast Cancer Detection :

We have studied and implemented various boosting and ensamble algorithms .After training these algorithms,they showed different accuracies based on their performance as follows:

- Support Vector Machine showed an accuracy of 96.49%.
- K-Nearest Neighbor showed an accuracy of 93.85%.
- Random Forest showed an accuracy of 97.36%.
- XGBoost Classifiershowed an accuracy of 98.24%.

Among the percentage accuracies obtained after training the algorithms, it is found that XGBoost Classifier-type of boosting algorithm showed highest percent of accuracy.

3.CONCLUSION

There are different Machine Learning techniques that can be used for the prediction of breast cancer to predict whether the tumor is benign or malignant. In this study Random forest, K Nearest Neighbors and SVM were used and the accuracy for the prediction on the dataset namely, Breast Cancer Wisconsin (Diagnostic) Data Set was found to be 97.6%, the highest of the three, using Random Forest. Further studies should be conducted to improve performance of these classification techniques by using more variables and choosing for a longer follow-up duration.

In this survey , the performance of different machine learning algorithms such as Support Vector Machine(SVM) and K-NN, XGBoost and Random Forest are assessed. Many researchers have applied the algorithm of neural networks for predicting cancers,especially the breast cancer. By going through various articles,XGBoost is applied for detecting breast cancer.

REFERENCES

[1] *BREAST CANCER DETECTION USING RANDOM FOREST, KNN AND SVM*RishabhKhosla Maharaja Agrasen Institute of Technology, New Delhi, India. *Vol.7*

[2] Early Detection of Breast Cancer Using Machine Learning Techniques M. Tahmooresi1 , A. Afshar2 , B. Bashari Rad1 , K. B. Nowshath1 and M. A. Bamiah2 1Asia Pacific University of Technology and Innovation (APU), Malaysia.-paper 3

[3] Breast cancer detection by leveraging Machine Learning Anji Reddy Vakaa , Badal Sonia , Sudheer Reddy K.b,∗ a Department of CSE, National Institute of Technology, Silchar, India b Researcher, Hyderabad, India-paper 2

[4] Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis Dana Bazazeh1 and RaedShubair 1,2 1Electrical & Computer Engineering Department, Khalifa University, UAE 2Research Laboratory of Electronics, Massachusetts Institute of Technology, USA