International Journal of Multidisciplinary and Current Educational Research (IJMCER)
ISSN: 2581-7027 ||Volume|| 3 ||Issue|| 4 ||Pages 115-121||2021||

(Breast Cancer Detection Using Machine Learning Classifier)

[1,]Gunja gayakwad, [2,]Chandni kumari
[1],*(http://orcid.org/0000-0003-1442-0324)*
[2],*(http://orcid.org/0000-0002-4488-9320)*

ABSTRACT: Breast cancer has become a common factor of today's . all general hospitals have the facilities for the confirmation of breast cancer through mammograms. Confirmation of the breast cancer for a long time may increase the risk of the breast cancer spreading. Therefore, A electronic breast cancer identification technology has been developed to reduce the time taken for the confirmation of the breast cancer and reduce the death rate. We have developed the technique which gives minimum error to increase accuracy. We have used different algorithm like SVM, Logistic Regression, Random Forest , KNN and Decision Tree to predict the breast cancer and to find better accuracy to detect the Breast Cancer . All experiments are supervise in JUPYTER platform .The future research can be carried out to predict the other different parameters.

KEYWORDS : Breast Cancer, Awareness, Risk factors machine learning, feature selection, classification, prediction, KN N , SVM, linear regression. This research paper focuses on breast cancer decision support system to explore ICA facility deficiency properties.. we have used WDBC which originally contains 30 features bi:t here one feature is reduced to make it a I-dimensional vector to compute on lC It is used to reach the diagnostic accuracy of the classifier like KNN, XGboost, linear regression, SVM. This classified are assessed how to efficiently differential tumors into benign In phrase of vagueness, responsiveness, perfection, F-score & venomous. This method reduces improved clinical decision support systems and making less computational complexity.

I. INTRODUCTION

The most generally cancer in women is breast cancer. Its number is increasing at a faster rate, According to an estimate. about 3 lakhs of new cases will be diagnosed in the '20s[1]. It is the leading source of decease in women, maybe due to its late diagnosis. the primary patients of breast cancer are adult women i.e. at age 45 or older[2]. A growing knowledge of the risk complexity of breast cancer has led to a development technique For early diagnosis of breast cancer. We are studying two types of breast cancer first one is the benign and the other one is malignant. Benign tumors are not cancerous hence not dangerous for life. But it can be the risk factor for breast cancer[3]. A malignant tumor is cancerous and dangerous than the benign one and is the major cause of mortality only early detection and the right diagnosis can solve the problem [4].

So with the aim of ameliorate the precision of classifications of breast cancer computer based diagnostic tools can be helpful types of machine learning tools which are discovered XGboost classifier ,KNN , Linear Regression and SVM[5] . this tools helps in differentiating the tumor as benign and malignant with excellent accuracy.. XGBoost Model for Classification XGBoost are stubby for supreme Gradient Boosting and is an efficient implementation of the non-linear gradient boosting machine learning algorithm. XGboost Classifier is a scikit-learn APIaccordant order for categorization[6].SVM (Support Vector Machine) an efficacious specifics learning way for categorization. SVM is established on discovering supreme hyperplane to disparate dissimilar types mapping input data into higher- dimensional attribute expansion[7]. The advantage of support vector machine is that it has fast training technique for large no. of data too. Hence it can be use for various identification difficulty for instance an gubbins identification and profile observation[8]. The purpose of this paper is to research the influence of attribute mitigation using ICA on classifying of the cancer just simultaneously non-malignant or malignant[9]. The original data set (WDBC) Wisconsin data breast cancer having 30 features is reduced into only one feature using 5 divided by 10 or ½ -fold fusion-validation and 20 percent subdividing to estimate the performance of *k*-NN, XGboost, Linear regression and SVM. Performance measures including vagueness, responsiveness, perfection, F-score & venomous to compare the classifiers[10].

Figure:- Stages of breast cancer

We are studying two types of breast cancer first one is the benign and the other one is malignant. Benign tumors are not cancerous hence not dangerous for life. But it can be the risk factor for breast cancer. A malignant tumor is cancerous and dangerous than the benign one and is the major cause of mortality only early detection and the right diagnosis can solve the problem. So with the aim of ameliorate the precision of classifications of breast cancer computer based diagnostic tools can be helpful types of machine learning tools which are discovered XGboost classifier ,KNN , Linear Regression and SVM . this tools helps in differentiating the tumor as benign and malignant with excellent accuracy and we studied from the research paperof Breast Cancer Classification and Prediction using Machine Learning – IJERTJean Sunny , Nikita Rane , Rucha Kanade, Sulochana Devi **(**03-03-2020**).** XGBoost Model for Classification XGBoost are stubby for supremeGradient Boosting and is an efficient implementation of the non-lineargradient boosting machine learning algorithm. XGboost Classifier is a scikit-learn APIaccordantorderforcategorization.

II. LITERATURE REVIEW

Recently machine learning algorithms were used in several medical data sets to categorize breast cancer. International meeting regarding cloud computing data science & engineering confluence in 2019 by Uma gupta and Savita goel proposed that main aim of the research is to predict the specificity of data mining algorithms to known chance of recurrence of disease in patients on basis of some basic parameters [11]. As per experiment, classification algorithm are better than clustering algorithm. The experiment concluded that the decision tree and SVM are the good predictor 81% accuracy on the clutch out sample. Whereas fuzzy c-means has lowest accuracy of 37%. For the detection of Breast tumor , machine learning algorithmic rule are applied on Wisconsin diagnostic breast cancer dataset via Abien Fred M Agarap [12]. here, six machine learning algorithmic rule are used for cancer detection they are -GRU -SUM, statistical regression , many layer perception, K-nearest neighbor search softmax downturn and support vector machine, decision tree, artificial neural network. These learning algorithms are applied on Wisconsin Diagnostic Breast Cancer (WDBC) dataset via evaluating their systemization test accuracy, and their responsiveness and specificness values for disease categorization. These dataset contains features computed from digit used image of FNA tests which were done on breast lump. The dataset is divided into 70% for training period and 30% for testing period for the implementation of ML algorithms. The result showed. Soaring production in binary categorization of carcinoma i.e. carcinoma or malignant. For detailed study, CV technique is used like K-fold cross validation technique. This provide better measures of model prediction performance & helps in predicting the format optimal parameters for machine learning algorithms tool[13].Evaluation of attribute assortment with categorization breast cancer dataset Indian j, computer sci. engg . (IJCSE) in 2011. lavanya D, rani D.K.U,[14]. proposed that which is published in 2011 in which WBC, WDBC, Breast Cancer dataset and decision tree with

and lacking feature choosing are used. Feature selection enhance the result is WBC: 96.99%, WDBC 94.77% , Breast Carcinoma 71.32% of this research paper. Yixuan Li and Zixuan Chen[15] used two datasets in the method.for cancer prediction known as production assessment of machine learning method. This study collects data from BCCD and WBCD. Which contains 116 volunteer plus 9 attributes and 699 volunteers + 11 attributes respectively. Then raw data of WBCD is processed to obtain info which hold 683 volunteers with 9 features to show the participant has malignant cancer. The accuracy of F – measures mark and ROC- curve of 5 categorization model are compared to obtain result that shows RF is primary model of this paper**.** The result of this research awards citation to discriminate the category of tumor. Some limitations of this project is that limited data has impact on accuracy of result so RF can be combined with other technologies to get solve the limitations and get efficient result in longer term work[16].

A comparative study is done by Mumine Kaya Keles [17]To predict & detect breast cancer speedy when the carcinoma is small, non-invasive and painless by using mining classification algorithms the subject of comparison includes data mining algorithms and weka tools. Here, weka data mining software is appealed to an antenna dataset to know the efficiency of data minng methods in detection of cancer. The dataset contains 6006 rows out of which 5405 are used as training dataset & rest 60 % are used as test data set. This dataset is converted to arff format is sed by weka tool in the form of file. It uses 10 fold cross validation with the help of knowledge production predicated on adaption learning data mining software implement to get optimum result. This gives average accuracy of 92.2%.

Haifeng Wang and Sang Won Yoon's [18] project known as breast cancer prophecy using data mining medium is used to test impact of feature space depletion. A essential integrant analysis method is used to bring down factor space. It contains hybrid between principal component analyses and related data mining replicas. Two test data set are used to measure the performance of these models. They are Wisconsin diagnostic breast cancer 1995. To estimate the test error of each model; 10 fold cross validation method is used. PCs-SVM has the highest for WBC data i.e. 97.47% and PCi -ANN is best in terms of accuracy for WDBC data with 99.63%. PCA(principal component analysis) provides better result because only principal component produce large pat of information, which can decrease data noise so that feature space is increased. "Machine Learning with solicitation in breast carcinoma Diagnosis and Prognosis" Is proposed by Wenbin Yue and Zidong Wang[19]. This provides ideas of ML approaches and their application in BC identifications and forecasting. ML techniques can boost the prediction accuracy & classification method. Dissimilar algorithms has distinct aspects and different mechanisms. Recently ML methods are applied in healthcare systems for intelligent work.[20]

III. BACKGROUND

Breast Cancer Classification:- classification of breast cancer mainly aims at the method of picking the dimples treatment. This classification divides carcinoma on the basis of their spread. lt is important as classification allow scientist to find, group and classify by a uniform system. Classification algorithms choose one or more discrete variables for classification[21].

there are two method is data processing they are classification and clustering. The purpose of the clustering method is to takeout information from the knowledge set to obtain clusters and to describe the information to set itself, while the purpose of classification is to classify unknown conditions supported by learning from existing patterns and information sets and is therefore future . make prediction, Conditions[22]. The rain set assesses the classifier. The most common algorithm used for cancer classification are support vector machine(SVM), KNN, linear regression, XGboost, Decision tree, AdaBoost, naive Bayes, random forest etc. Researchers try to find the simplest algorithm to understand the exact classification result. But variable data Quality affects results .Earlier detection provides a better option for treatment and hence better chance of survival. Regular checkup early detection can save women's life[23].

Machine Learning Algorithm:- This algorithm is an application that empowers systems to learn and improve from experience without being programmed manually. Machine learning depends on the occurrence of the computer program that will use to access the data and to use them [24].

1. **Naive Bayes:-** It is a classification machine learning technique. The notion of freedom among prophets. A naive base.
2. In this, the classifier assumes that the attribute in one class is not related to the presence of another particular feature.

3. **Random forest:-** The lieutenant makes a large number. Of trees that achieve their output through learning method for regression classification. In order to construct the tree, it uses bagging and feature randomness. It does not overfit the data
4. Support Vector Machine: -. It is used as a training algorithm to study classification and regression rules from data. SVM is used when the number of features and the number of instances are high..
5. K nearest neighbors: K-Nearest Neighbor is the most usable machine learning algorithm because the data given in it is labeled. This is an unconventional method because the classification of a test data point depends on the closest training data points, rather than considering the dimensions (parameters) of the dataset.
6. **Decision Tree:** Decision tree algorithm is a learning algorithm. It can also be used for regression and classification. The goal of using a decision tree is to create a training model that can be used to estimate the class or value of a target variable to someone who learns from prior training data by learning general decision rules.
7. **Adaboost:** - AdaBoost, short for Adaptive Boosting, is a machine learning meta-algorithm originated by Yoav Freund and Robert Shepire, who won the 2003 Gidel Prize for their work. LT can be used in concurrencewith many types learning algorithms to better performance.
8. **XGboost:-** XGboost 'is another machine learning algorithm which is an open-source library. It provides under grading boosting methods.

Proposed Model : To predict breast cancer patients, we have given a classification model to enhance accuracy. The framework consists of the on account of major steps Dataset Selection, Data processing, Research by exhibit (Training) i.e. SVM (support vector machine), statistical Regression and KNN, Achieving inculcated model with high-rise correctness, Using the constructed model for forecasting. In this objective model we compare various machine learning (ML) algorithmic rule: such as Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RT), Adaboost classifier, Naive Bayes (NB), K-Nearest Neighbour (KNN) search, XGboost classifier linear regression. We acquired the data set from the Wisconsin dataset. We used that data set for the execution of the ML algorithms, the dataset is divided into two parts, the first one is the training set and the another one is the test set.After training process, The use of test data defines the diagnostic performance of the classifier in phrase of responsiveness, explicitness , accuracy, f1- score. The algorithm yielding the best results will be supplied as a model of dominance. The classifier predicted malicious prediction or made a benign prediction. The data set is obtainable at the UCI Machine Learning Fund. It has 32 real-world features. . The process of the proposed system is as follows, and finds the highest accuracy

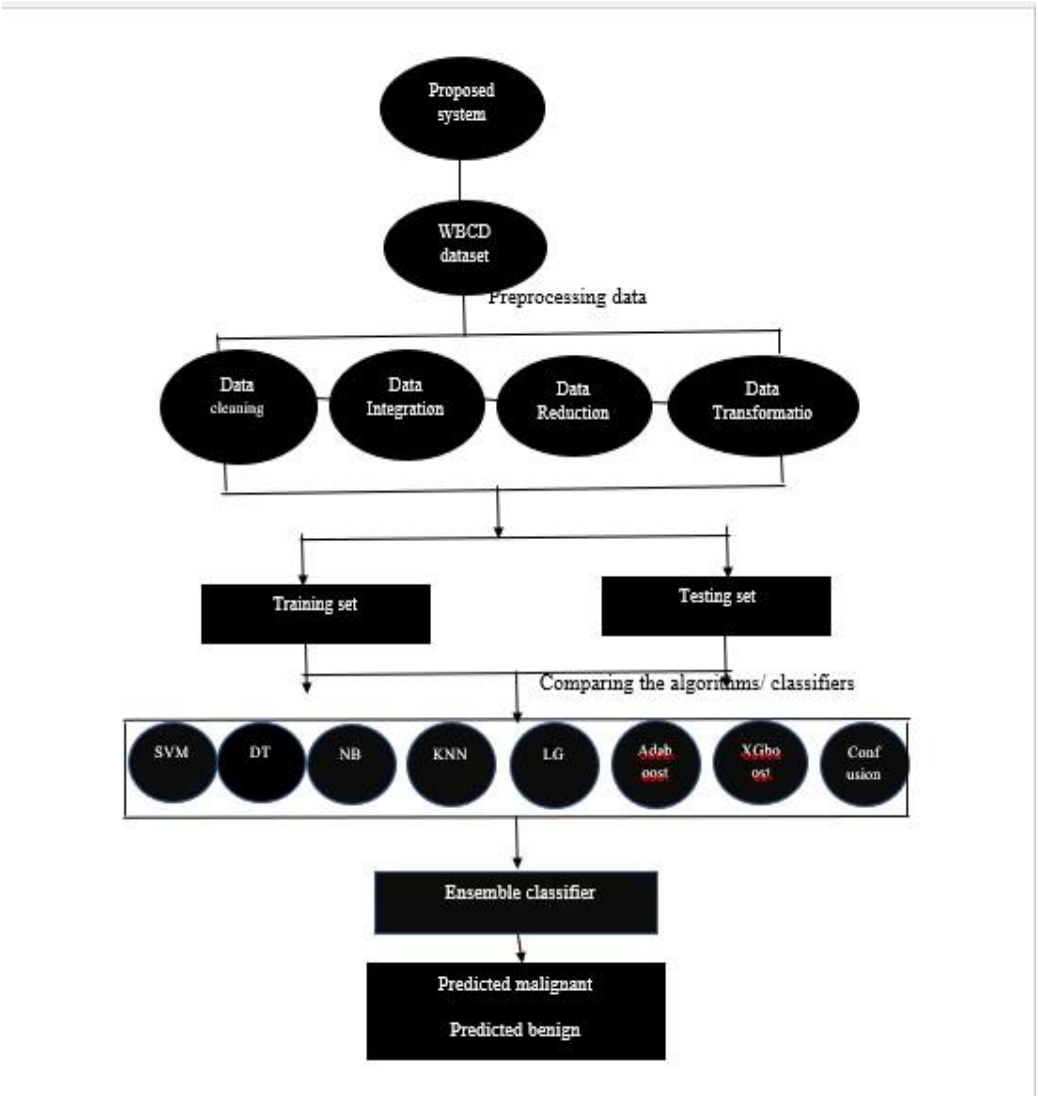

Fig:- Block Diagram

IV. RESULTS

Out of 600 patients, 502 patients and 98 patients be alive and life less, respectively. In this research paper, KNN, SVM, and XGboost technique displayed better results in correspondence to the other techniques (NB, AD, SVM and XGboost, linear regression , Decision tree, confusion matrix). The accuracy, sensitivity, and the F1 score are 94%, 96%, 98%, respectively. However, the decision tree machine learning strategy come up with poor performance (accuracy 75%, sensitivity 76%, and F1-score 0.98) . underlying, the eight categorization algorithmic rule were tested on the WDBC datasets without appealing the pre-processing techniques. Inter alia , the supreme result was declared for J48: 75.52% in the Breast carcinoma dataset and for SMO: 96.99% in the WBC dataset. Upcoming, in the wake of applying pre -processing capability exactness increases to 98.20% with XGboost in the Wisconsin Breast Carcinoma dataset and 97.56% with SVM in the WDBC dataset.

V. CONCLUSION

chine learning perhaps a very good assist in determining the line of investigation to be escort by extorting knowledge from such suitable databases. In this evaluation, we have attempted to · illuminate and evaluate, different machine learning's performance which is used to predict and prognosis of cancer. In that research paper, we concerned how to deal with variance data that have lost utility using retesting techniques to enhance the categorization accuracy of recognizing breast carcinoma. In our research , eight classifiers algorithms SVM, KNN, DT,Confusion matrix, linear regression, adaboost, xgboost , NB, and random fosrest applied on wisconsin breast cancer datasets. Outcome display that using the explore strainer in the pre-

processing period increase the classifier's presentation. In the upcoming, the unchanging examination will try to dissimilar classifiers and distinct datasets.

REFERENCE

[1] International meeting regarding cloud computing data science & engineering confluence in 2019 by Uma gupta and Savita goel proposed that main aim of the research is to predict the specificity of data mining algorithms to known chance of recurrence of disease in patients on basis of some basic parameters .

[2]. On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset Abien Fred M. Agarap On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset (arxiv.org)

[3]. WBCD breast cancer database classification applying artificial metaplasticity neural network Author links open overlay panelA.Marcano-CedeñoJ.Quintanilla-DomínguezD.Andina .

[4]. Lavanya, D., Rani, D.K.U.: Analysis of feature selection with classification: breast cancer datasets. Indian J. Comput. Sci. Eng. (IJCSE), pp. 756–763 (2011)

[5]. Discrimination of breast cancer from benign tumours using Raman spectroscopy Fiona M. Lyng ,Damien Traynor,Thi Nguyet Que Nguyen,Aidan D. Meade, Fazle Rakib,Rafif Al-Saady, Erik Goormaghtigh, Khalid Al-Saad, Mohamed H. Ali Published: February 14, 2019

[6]. Performance Evaluation of Machine Learning Methods for Breast Cancer Prediction Applied and Computational Mathematics AuthorsYixuan Li, School of Mathematics and Statistics, University of Sheffield, Sheffield, UK Zixuan Chen, School of Information, Zhejiang University of Finance and Economics, Hangzhou, China Volume 7, Issue 4, August 2018, Pages: 212-216

[7]. Breast cancer prediction and detection using data mining classification algorithms: A comparative study MK Keleş Tehnički vjesnik 26 (1), 149-155 [8]. Proceedings of the 2015 Industrial and Systems Engineering Research Conference Project: Machine Learning and Data Mining TechniquesAuthors: Haifeng WangMississippi State University,Sang Won YoonState University of New York.

[9]. A PSO-based deep learning approach to classifying patients from emergency departments March 2021International Journal of Machine Learning and Cybernetics Authors: Weibo Liu, Zidong WangHarbin University of Science and Technology ,Nianyin Zeng, Xiamen University, Fuad E. Alsaadi

[10]. Rodrigues, B.L.: Analysis of the Wisconsin Breast Cancer dataset and machine learning for breast cancer detection. In: Proceedings of XI Workshop de Visão Computational, pp. 15–19 (2015)

[11]. Asri, H., Mousannif, H., Al, M.H., Noel, T.: Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Comput. Sci. 83, 1064–1069 (2016)

[12]. Silva, J., Lezama, O.B.P., Varela, N., Borrero, L.A.: Integration of data mining classification techniques and ensemble learning for predicting the type of breast cancer recurrence. In: Miani, R., Camargos, L., Zarpelão, B., Rosas, E., Pasquini, R. (eds.) GPC 2019. LNCS, vol. 11484, pp. 18–30. Springer, Cham (2019).

[13]. Analysis of Breast Cancer Detection Using Different Machine Learning Techniques Author: Siham A. Mohammed , Sadeq Darra (July 2020) [14].Machine Learning: Algorithms, Real-World Applications and Research Directions Iqbal H Sarker [15].Machine Learning: Proceedings of the Thirteenth International Conference, 1996. Experiments with a New Boosting Algorithm Yoav Freund Robert E. Schapire. [16]. Breast Cancer Research and Treatment William J. Gradishar Breast Cancer Research and Treatment | Home (springer.com)

[17]. High-resolution integrated map encompassing the breast cancer loss of heterozygosity region on human chromosome 16q22.1 E Frengen 1 , P Rocca-Serra, S Shaposhnikov, L Taine, J Thorsen, C Bepoldin, M Krekling, D Lafon, K K Aas, A A El Monéim, H Johansen, M Longy, H Prydz, F Dorion-Bonnet. [18]. Survey of Data Mining Techniques for Prediction of Breast Cancer Recurrence Desta MulatuRupali R.Gangarde(2017) http://www.ijcsit.com/docs/Volume%208/vol8issue6/ijcsit201708060

[19]. Breast cancer: Introduction October 2001 Seminars in Cancer Biology 11(5):323–326 DOI: 10.1006/scbi.2001.0381 Project: Molecular biology and genetics of breast cancer Authors: Sigurdur Ingvarsson University of Iceland (PDF) Breast cancer: Introduction (researchgate.net)

[20]. On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset Abien Fred M. Agarap On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset (arxiv.org) [21] WBCD breast cancer database classification applying artificial metaplasticity neural networkAuthorlinks open overlay panelA.Marcano-CedeñoJ.QuintanillaDomínguezD.Andina.

[22]. Implementation of machine learning algorithms for different datasets using python programming language. Renuka sagar, Dr U Eranna http://www.ijser.org/researchpaper/Implementation-of-machine-learning-algorithms-fordifferent-datasets-using-python-programming-language.pdf.

[23]. LncRNA PVT1 promotes malignant progression in squamous cell carcinoma of the head and neck Changyun Yu, Yunyun Wang, Guo Li, Li She, Diekuo Zhang, Xiyu Chen, Xin Zhang, Zhaobing Qin1, Hua Cao1,Yong Liu. [24]. Practical Bayesian Optimization of Machine Learning Algorithms June 2012 Advances in Neural Information Processing Systems Source arXiv Authors: Jasper Snoek Hugo Larochelle Université de Sherbrooke Ryan P. Adams.