ISSN No:-2456-2165

Mammary Neoplasm Prognosis using Machine Learning: A State of the Art Survey

Annapoorna B. R., Kota V. Vishnu, Reena Jasmine Edwin, S Sai Brinda, Shalini Singh
Department of Computer Science
Dayananda Sagar College of Engineering Bangalore, India

Abstract—Breast cancer is the most frequent cancer in women worldwide and is also the most lethal. The reasons for this illness are many and challenging to identify. Furthermore, the diagnostic technique, which determines whether the cancer is benign or malignant, requires substantial work from doctors and physicians. There are many diagnostic tests possible which can be conducted by medical professionals to detect it; however, it has been increasingly strenuous to precisely spot and acts on its prognosis As a result, in recent years, there has been a surge in the use of machine learning and Artificial Intelligence in general as diagnostic tools. ML seeks to make computer selflearning easier. in lieu of contingent on explicit pre-programmed rules and models, it is based on finding patterns in observed data and creating models to predict outcomes and evaluate them on performance measure features like accuracy, precision, and recall. The primary impetus of this review is to culminate all the antecedent studies of machine learning algorithms being utilised for breast cancer prediction. this survey is going to be useful to the researchers because of the elaborated probe of various methodologies for undergoing supplemental inquisitions.

Keywords:- *Breast Cancer, Medical Diagnosis, Machine Learning, Logistic Regression, KNN, Decision Tree, SVM, Random Forest, Naive Bayes.*

I. INTRODUCTION

Cancer may be a generic term for an immense cluster of diseases that may affect any body part. alternative terms used are malignant tumours and neoplasms. One process feature of cancer is the fast creation of aberrant cells that grow on the far side of their usual extremities and which may then overrun contiguous elements of the body and unfold to alternative organs; the latter method is stated as metastasis. Widespread metastases are the first reason for death from cancer. Breast cancer accounts for over a quarter of the total fatalities precipitated by all forms of cancers worldwide. When cells in the breast start to grow out of control, cancer develops. Tumors are these collections of cells which can be detected on a Radiography scan or felt as a lump. When tumorous cells are introduced inside the bloodstream of an individual, cancers are prone to be dispersed. The lymph nodes are the outlet for which blood is transported to all areas of the body, and any intrusion is usually caused through the same. The most ubiquitous kinds of breast cancer are ductal carcinoma in situ (DCIS) and invasive carcinoma. Tumors come in two varieties. One is benign, meaning it isn't cancerous, and the other is malignant, meaning it is.

In modern society, early sickness detection is essential. As the population increases, there is an exponential increase in the likelihood of dying from breast cancer. Breast cancer is the most common, ubiquitous and most deadly cancer in women. There are several, elusive causes for this condition. Furthermore, it takes a lot of effort on the part of doctors and medical professionals to perform the diagnostic procedure that establishes whether the cancer is benign or malignant.

When many tests, such as homogeneity and clump thickness and cell uniformity, is used to determine its potency. As a result, the use of machine learning and all forms of artificial intelligence as diagnostic tools has increased recently. ML aims to simplify computer self-learning. It is based on identifying patterns in inspected data and building models to anticipate outcomes and evaluating them on performance measure aspects like accuracy, precision, and recall rather than relying on explicit pre-programmed rules and models.

This study aims to provide a variety of methodologies for studying the usage of various machine learning (ML)based algorithms for early breast cancer diagnosis using the Wisconsin breast cancer dataset. We'll look at and contrast the machine learning methods used for classification, including KNN, SVM, LR, NB, RF, and Decision Tree (DT), for computing accuracy in terms of performance metrics such as recall, accuracy percentage and precision F1 score.

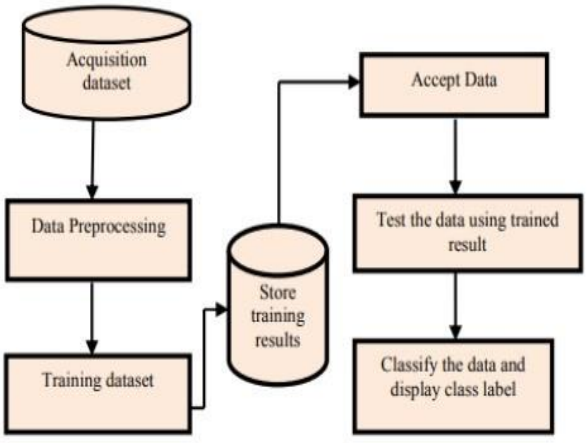

Fig. 1: Approach Overview

ISSN No:-2456-2165

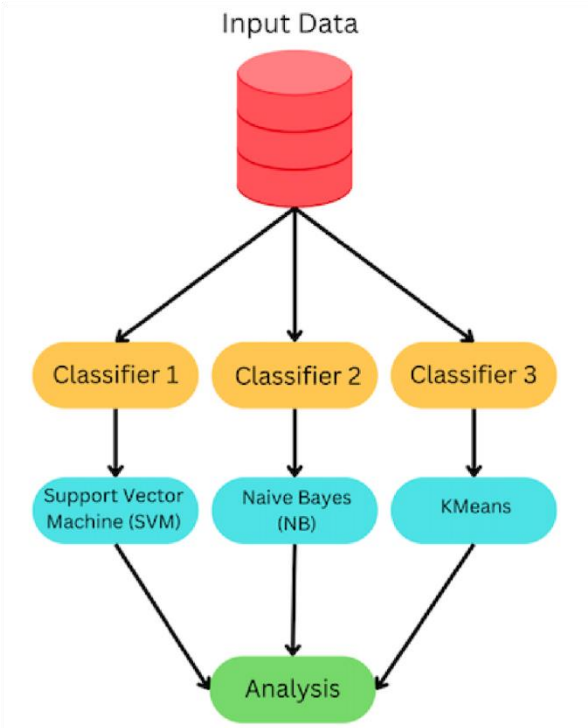

Fig. 2: Proposed System

II. LITERATURE SURVEY

The authors[1] have used VGG16 and Resnet50 models to classify between normal and abnormal cancer tumours. Data taken from the IRMA dataset is processed and resized before getting evaluated. The CNN consisting of several layers is used to pool,flatten and sample test cases to test contours and ridges formed. VGG16 is used when high computational requirements are needed whilst Resnet is used for skip connection to pass input to the subsequent layers of the model.

The VGG16 and Resnet provide an accuracy and precision F1 score of over 85 percent.

The authors[2] have utilized Naive Bayes Algorithm Classifiers to segregate the data with minimum amount of training required. The data is classified based on it‘s class, variable and attribute name. The correspondence to each supposition in the data is allocated with respect to Gaussian distribution with mean and standard deviation analyzed. The data collected from this analysis is subjected to confusion matrices to check for True Positive,True Negative,False Positive,False Negative values to measure accuracy and F1 Scores. Naive Bayes Algorithm produces a score of 98 percent using this method.

The authors[3] use WEKA to analyze data from the UCL ML repository to predict the condition of Breast Cancer in a tumour scan. Several factors like cell size, shape , nucleoli, Clump Thickness, Marginal Adhesion is considered before coming to a conclusion. Naive Bayes uses Gaussian Distribution to cluster the data based on the results obtained. Using WEKA along with Naive Bayes yields an accuracy of 94.08 percent after segmenting the results into benign and malignant clusters.

The authors use Naive Bayes Classifier algorithms to analyze and sort Image Mammography scan results to Proportional k-Interval Discretization (PKID) and DISCRETIZE filters. PKID Filters allow doctors to sort the scan results based on the data received from the tests conducted after sorting through itemised test cases. A confusion matrix is used to classify the results into benign and malignant which gives an accurate measure of the TP,TN,FP,FN .. This method gives an accuracy and F1 score of 74 percent.

The authors have used various algorithms like Randomn Forest, kNN (k-Nearest-Neighbor) and Naive Bayes to distinguish between instances and attributes in various Mammogram scans. Linear Discriminant Analysis is used for feature selection to train the model using fuzzy interference. The algorithms used to train and test the model are divided into supervised and unsupervised methods and ranked according to factors like Time Complexity and Model Parameters. Random Forest , Naive Bayes and kNN algorithms all display accuracy and F1 scored greater than 91 percent using this method.

The aim of the authors[6] is to develop an early-stage breast cancer detection system capable of automatically classifying irregularities in mammography images acquired from the Mammographic Image Analysis Society(MIAS) database. First the data preprocessing is done using a Median Filter to remove noise, further it is segmented using the OTSU thresholding technique. GLCM, Second Order Texture, is used to extract features.The machine learning algorithm employed is K-Nearest Neighbor.The accuracy score according to this algorithm is 92 percent.The authors of the research want to increase classification accuracy by employing additional best classification methods.

The research work provides in-depth analyses of the technical and usability aspects of histopathological image characteristics and performs breast cancer diagnosis using the Breakhis and breast histopathology image datasets.A wellstructured dataset is generated by repeatedly extracting 13 Haralick texture characteristics from each histopathology image.The dataset generated is subjected to dimension reduction techniques like PCA and LDA. The machine learning technique used to identify breast cancer is K-Nearest Neighbor Classifier. Accuracy score of KNN using LDA was 80.0 percent,which was higher than the accuracy score of KNN using PCA, which was 56.0 percent.Whenever a dataset has texture features, the approaches suggested by authors[7] may be used to get insights into which factors contribute the most to the target features.

The authors[8] studies the application of parallel programming for breast cancer classification and prediction on large datasets such as the Wisconsin Breast Cancer dataset.The dataset is used to compute the kNN method both serially and parallelly. The findings are validated by employing frameworks such as Compute Unified Device

ISSN No:-2456-2165

Architecture (CUDA) and Message Passing Interface (MPI),which divide the workload and thus finish the operation in considerably less time. The results highlighted that parallel execution consumes less time (almost half) than sequential execution. In the future, authors may plan and put into practice to operate in a parallel setting with less communication overhead.

The authors[9] proposed an improved strategy that eliminates the need for the K value in the KNN algorithm used to diagnose breast cancer while maintaining the same performance and enhancing it for particular datasets since this value affects the algorithm output performance. Instead of employing K values in the KNN algorithm, this technique employs a ”zone classifier” approach to classification and accuracy maximization. The goal is to build a classification zone for each new data point and apply the most often occurring class in the defined zone to this new element. The Zone Classifier technique yields an Accuracy score of 93 percent, Precision score of 96.5 percent, and Recall score of 94.74 percent. This strategy eliminates the overhead of the user selecting a K value to provide as input and yields a better result. The proposed method was applied for small and medium datasets.In future research, authors intend to improve technique for application on large datasets.

The classification and prediction algorithms are examined by utilizing information from wisconsin university database to ascertain if breast cancer is benign or malignant utilising data mining techniques.The knn prediction algorithm and classification algorithm are employed in this paper. Since an .ARFF file is used as the application’s input ,the file conversion is performed using the Weka interface 3.6 and the experiment was conducted in matlab(matrix laboratory). the knn prediction as well as the knn classification algorithms had success rates of 80-90 percent and error rates of 10-15 percent.

There is still potential for improvement in the algorithms so the authors[10] can handle more input combinations and increase success rates.

The research work describes Naive Bayes improved KNearest Neighbor technique (NBKNN) for diagnosing breast cancer on the UCI repository. Data cleaning is performed on the input data to eliminate outliers and missing values. Each classifier’s effectiveness is evaluated using a 10-fold crossvalidation method.Traditional classifiers such as KNN and naive Bayes, as well as the NBKNN algorithm, are employed to predict cancer and compare the findings.The accuracy score achieved by KNN is 96.7 percent, the accuracy score acquired by Naive Bayes is 95.9 percent, and the accuracy score obtained by NBKNN is 97.5 percent. The study shows that the NBKNN approach outperforms other conventional classifiers in terms of performance and that several techniques influenced by Biology such as Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) may be utilized to optimise the results.

The motive of authors[12] is to use decision tree algorithm to develop a classification model for breast cancer with the added notion of preprocessing the input data without eliminating the missing values.Breast Cancer data from UCI Machine Learning was used. The decision tree model is applied to two sets of data, the first being the most often used strategy, in which incomplete readings are eliminated, and the second being missing readings that are replaced with mode for each attribute.The Decision Tree technique uses the GINI index approach to select the parameter as the root node.Data with missing readings deleted obtained an accuracy of 85.2 percent and Data with missing readings replaced with mode of data obtained an accuracy of 78.57 percent.

The authors[13] proposed a breast cancer detection model using microarray breast cancer gene expression data. A hybrid of two choice of feature selection techniques: the filter method using Fisher-score and the C5.0 algorithm’s inner feature selection capability are applied.This is employed because the most prevalent issue with data on gene expression is its high dimensionality.Support vector machines, C5.0 Decision Trees, Logistic Regression, and artificial neural networks are the classification methods that were employed to evaluate the predictive accuracy of this strategy. Prior to the application of feature selection, 24481 genes were chosen, with ANN showing a better accuracy score of 86.99 percent and C5.0 showing the lowest accuracy score of 79.01 percent. When feature selection is applied,the number of genes chosen was reduced to 5 and all shrinkage models provided classification accuracy greater than 80 percent. The authors intend to examine the effectiveness of the suggested strategy using new datasets from microarrays that have varied qualities that differ in the quantity of classes, genes, and samples.

On the Wisconsin breast cancer dataset, the authors[14] extensively analysed the predicted execution of SVM, Gaussian Naive Bayes, K-Nearest Neighbors, and Classification and Regression trees (CART). Various classification algorithms’ classification accuracy, precision, and F-measure are investigated. The experiment shows that the SVM classifier is a superior choice for classifying since the algorithm’s performance is enhanced by tuning the dataset’s parameters and reduces the likelihood of overfitting. The optimal criteria, on the other hand, are required for accurate categorization.

For breast cancer prediction, the authors[15] used two strong classifiers, ANN and DT . Experimental findings reveal that these approaches have potent effects for this task, with the general prediction accuracy of the DT from ninety to ninetyfour percentage. furthermore, Support Vector Machine ranging from 94.5% to 97%. It is observed that DTs need a little preparatory work. However, when there are numerous consequences, they become unsteady and challenging to decipher.

ISSN No:-2456-2165

Yixuan Li et. al.[16] employed the LR, DT, RF,SVM and NN models to prognosticate the kind of breast cancer with other features. The prediction findings will aid in lowering the rate of false - positive results and developing appropriate therapeutic plans for recovery. In this investigation, 2 datasets are employed. This analysis initially gathers source data from the BCCD dataset, that has 116 participants along with nine characteristics, and source data from the WBCD dataset, which comprises 699 participants containing 11 features. The source data from the WBCD dataset was then preprocessed, yielding data including 683 participants with nine characteristics and an index signifying whether the volunteer had a malignant tumour. Off the back of collating the accuracy, The ROC curve and F-measure metric of five different classifiers were used to determine which model should be used as the principal classifier in this investigation. It performs well on huge datasets. They are, however, significantly more difficult and time-consuming to build. This experiment only analyzes the data on 10 features. The lack of source data has an impact on the correctness of the outcomes. Furthermore, the RF may be used in conjunction with other approaches to data mining to provide more precise diagnostic conclusions.

The authors[17] of this research tested the accuracy, precision, sensitivity, and specificity of each algorithm: kNN,SVM,C4.5, and NB on the Wisconsin Breast Cancer datasets. According to the experimental data, SVM provides the best accuracy of 97.13% with the minimal false positive rate. Because every single trial is executed in a simulated conditions and with the WEKA data mining tool, the risk of overfitting is reduced. The best parameters, on the other hand, are required for proper categorisation.

The authors[18] have employed five primary algorithms: Random Forests, SVM, K-NN,Logistic Regression, Decision Tree to compute, contrast and assess various findings attained elicited from sensitivity, confusion matrix, AUC, accuracy, and precision to discover the superlative machine learning algorithm that is exact, dependable, and finds the highest accuracy. In the Anaconda environment, all algorithms were written in Python using the scikit-learn module. After a thorough evaluation of the models, it was discovered that the Support Vector Machine outperformed all other methods in terms of efficiency (97.2%), precision (97.5%), and AUC (96.6%). However, To achieve greater accuracy, new parameters are can be used for larger sets of data with more illness types.

The authors[19] used Grid search to present a model for predicting breast cancer using Support Vector Machine. The Initial Support Vector Machine model is evaluated in the absence of grid search. The Support vector machine model is then evaluated using grid search. Ultimately, a comparison study was performed, and a new model was created based on the results. The new model uses a grid search of data prior to fitting it for classification, which optimizes the outcome and produces much improved outcome than a conventional SVM model. It can be observed that the correct parameter values for gamma and C are crucial for a certain quantity of data. This approach could also be employed to anticipate other ailments, acting as a decision-support system in the healthcare division.

The authors[20] have led a series of investigations using machine learning models to enhance breast cancer categorization for the data set. It was demonstrated that logistic regression when implemented to training set, yields good findings. Utilizing seaborn and sklearn metrics, the accuracy is assessed and the confusion matrix was illustrated. This model achieves an accuracy of 97.63% . However, incremented data may be included in the data set, and accuracy can be boosted.

The authors[21] used Logistic Regression for Breast Cancer Detection. It was observed that the logistic regression method had an accuracy of more than 94% in detecting whether the cancer was malignant or benign.The findings indicate that integrating multidimensional data with various categorization, feature selection, and dimension reduction strategies might give beneficial tools for analysis in this domain. More research is necessary to enhance the efficiency of classification systems so they are capable of predict additional variables.

To identify the area of concern and identify anomalies in mammogram images, the author[22] employed a CAD system. The method employed in the paper is to locate and categorize tumors using digital mammograms. the data had been preprocessed using a Gaussian filter and an adaptive histogram equalization approach for smoothening of image and improving contrast. Otsu's approach is used for segmentation for extorting malignant tumors. Features are extracted using GLCM. The best features that will increase the efficiency of the algorithm are chosen using FCBF. In order to classify, RF is used as the classifier. For Evaluation of the result F measure, Confusion Matrix and ROC curve are used. The result has an accuracy 97.32%. By lowering the FP and FN, the result demonstrates that RF classifier enhances classification. This is valuable for radiologists in detecting malignant tumors in digital mammograms.

The proposed method in the paper[23] is Hierarchical Clustering Random Forest (HCRF) and Variable Importance Measure(VIM), for classification and feature selection based on the Gini Index respectively. The parameters of our model are selected using the grid search algorithm. Datasets utilized for the study include WBC and Wisconsin Diagnosis Breast Cancer. From the specified training set, several different training subsets are created using the bootstrap sampling technique. The trees that share similarities are grouped, together. In the end, we choose the decision tree from each cluster that has the highest area under the curve and discard the others. The developed model performs better when tried to compare to other classifiers like Adaboost and decision tree. The Selected Tree for Random Forest are of low similarity.On the WDBC dataset, our suggested technique achieves an accuracy of 97.05 %, and on the WBC dataset, it achieves an accuracy of 97.76 %.

ISSN No:-2456-2165

The author of the paper[24] has employed a variety of classification techniques, including LR, DT, and Random forest. The UCI Machine Learning Repository provided the Data Set (Wisconsin Breast Cancer). The goal of the research is to specify whether a tumor is benign or malignant and whether or not it is curable at that time. The dependent, independent, and other qualities are regressed with one another in the case of linear regression, and the resulting result has a success of 84.15%. All observation is provided to the decision tree and the most common outcome is considered as output. It takes the most opted option for all classification models. RF classification has a success rate of 88.14%.

The research[25] compares the results using two algorithms: Random Forest and XGBoost. UCI Machine Learning Repository is where information was gathered from. Trimmed means and modes had been employed for data preprocessing. This removes the data with extremely high and low values. 37% of the data is used for testing and 67% for training. The two aforementioned algorithms were then each applied separately. For the subset, RF employs bagging, and accuracy rises with tree count. The gradient boosting framework uses an algorithm called XGBoost that is based on decision trees. Calculations have been made for the F1 score, precision, the test's ability to correctly identify patients without cancer, as well as the test's ability to correctly identify patients with cancer. For Random Forest, their accuracy was 74.73%, while for XGBoost, it was 73.63%.

In the proposed model author suggests a method for locating Micro calcifications, tiny calcium apatite crystals that, despite their tiny size and low contrast, are the first indication of breast cancer. A coded contour is available with an image containing microcalcification denoting the area of their presence. An automated method employing discrete wavelet transform for segmenting and RF for classifying breast microcalcifications in mammograms respectively. The Digital Database for Screening Mammography has 966 mammography images divided into three classes: benign, malignant, and normal. To enhance, mammography images were processed through a two-dimensional discrete wavelet transform. The tissue surrounding the microcalcification is removed using the maximum entropy approach. The sequential forward features selection procedure is used to minimize the set of features after the features are chosen using GLCM.

Following that, Random Forest is used and a grid search was used to determine the parameters. It was trained using 10-fold cross-validation. In comparison to previous models, this one has a 95% accuracy rate.

For classification in the paper, we used the Sklearn Library's logistic regression. In a short amount of training time, this approach offers a strong prediction. The WBC Data Set serves as the source of the data set. Malignancy is directly correlated with tumor size and texture, according to the dispersed plot. With mean radius and texture, logistic regression is 90.48% accurate, whereas it is 96.5% accurate with maximum texture and maximum radius. Utilizing machine learning is a quick method of detecting cancer.

The Random Forest Algorithm has been used in the paper[28] to classify breast cancer cases. Here, the characteristics of various Eigenvalues and the output of different decision trees have been combined using Random Forest to increase accuracy. Sampling is done by bagging, and then we create decision trees using the CART algorithm. When separating nodes, they employ the Gini coefficient technique. The result that trees produce and the outcomes that they deliver after being trained are independent of one another. To provide results, many week classifiers are combined. It has an accuracy of 95 %.

The dataset used in the paper[29] was sourced from the Kuppuswamy Naidu Hospital in Coimbatore, India. Data is gathered from hospital charts, pathology reports, and other sources. To identify and categorize breast cancer, the paper used Expectation Maximization (EM) Based Logistic Regression (LR). Based on variables including family history of breast cancer, nipple level, lump location and size, breast nipple position, menstrual cycle, normal habits, number of miscarriages, diet, menopause, feeding, basic health hygiene, etc., the 82 cancer patients are studied and sorted. Beginning with the conversion of metadata into data, the EM based logistic Regression Result is used to compare different TNM stages by Chi Square Test method. Missed Classification Measures,Perfect Classification Measures, False Alarm Measures, etc. are the benchmark metrics taken into consideration here. The result of EM-based logistic regression showed average accuracy of 92 %.

The authors use Recursive Elimination, Unvariant Selection, Univarient Selection to present data processing results on various phase classifiers. Phase 0 uses attributes like Family tree, Breast Feeding, OCP, Axillary lymph node status, Ultrasonography (USG), Mammogram, True Cut Biopsy, Biopsy, HER2 status, ER, Diagnosis which gives features on patient details. Phase 1 uses Gaussian DB and sci-kit modules on BCDS.The data set (BCDS) is organized into attributes and class label when used in the classifier. Phase 1 is checked using measures like TN,FN,TP,FP. Phase 2 uses a mathematical function called Chi-Squared which is a statistical test to extract important features. Recursive Function is used to remove the lowest ranked features and new prominent contours are obtained. Phase 3 uses both Uni-variate Selection method and Recursive Elimination method to extract important documentations in order to have a higher relevance value. The data is then fed to the Naive Bayes Gaussian model to obtain a comparative study. This method reaches an Accuracy and Precision score of over 84 percent and have a computational run-time of around 10 seconds.

The authors use intelligent ensemble techniques like SMO, RF and iBK.Five individual classifiers Naive Bayes, SVM, Simple Logistics, Random Forest and iBK. The classifier algorithms perform better when multiple algorithms are used in a singular simulation compared to seperate simulations. The paper introduces a classifier mix-up using WEKA and BCDA. Combining input classifiers

with Logistics brings out a holistic approach to process and evaluate the data.Integrating several classifiers boosts the performance and the foundational capabilities of the model for future calculations. Stacking ensemble algorithms together give SMO the highest accuracy of 83 percent.

The authors propose using the firefly algorithm to decrease the variances in breast cancer mammogram scans.The algorithms performance is tested against known parameters like accuracy, specificity, sensitivity, and MCC which reveals that it produces better results than DWPT, SVM and BMC. CAD Software is used which seperates the scans into CT and MRI results. The scans after being classified into benign and malignant can be sorted and used to perform deliberations Wavelet packet techniques are used to overcome the flaws persent in the firefly ensemble algorithm. The algorithm is tested on two datasets which are MIAS and DDSM. The authors use CNN features like flattening and dropping to understand the depth perception of the obtained model . The CNN model is passed through a feature matrix which and divide the sample set into scaffolds. The firefly algorithm is then used to the run test cases on feature value and parameters like Neighbours. Several features like Sensitivity, Specificity and Accuracy can be derived after processing. It produces an accuracy of over 85 percent.

III. CONCLUSION

This study sets our major aim to discover the best suitable algorithm that can predict the occurrences of breast cancer more effectively by reviewing several machine learning algorithms for the prediction of breast cancer. The primary goal of this review is to highlight all previous studies of machine learning algorithms used for breast cancer prediction, such as (NB), (KNN), and (LR), as well as (RF), (SVM), and (DT), for computing accuracy in terms of performance metrics such as recall, precision F1 score, and accuracy percentage.

This research investigates how the usage of Neural Networks might effect changes in reports resulting from the examination of diverse datasets. When applied to a Neural Network, ensemble approaches provide a comprehensive view of the data and make it more intelligible.

REFERENCES

[1.] Nur Syami, Ismael, C. Sovuthy ”Breast Cancer Detection Using Deep Learning” doi:10.1109/EnCon.2019.8861256
[2.] Hajer Kamel, Dhahir Abdulah,Jamal M.Al-Tuwaijari ”Cancer Classification Using Gaussian Naive Bayes Algorithm” doi:10.1109/IEC47844.2019.8950650
[3.] Megha Rathi, Arun Kumar Singh ”Breast Cancer Prediction using Na¨ıve Bayes Classifier” International Journal of Information Technology Systems, Vol. 1; No. 2: ISSN: 22779825 (July-Dec. 2012)
[4.] Tawseef Ayoub Shaikh , Rashid Ali ”A CAD Tool for Breast Cancer Prediction using Naive Bayes Classifier” doi:10.1109/ESCI48226.2020.9167568
[5.] Shubham Sharma, Archit Aggarwal, Tanupriya Choudhury ”Breast Cancer Detection Using Machine Learning Algorithms ” doi:10.1109/CTEMS.2018.8769187
[6.] Than Than Htay,Su Su Maung, ”Early Stage Breast Cancer Detection System using GLCM Feature extraction and K-nearest Neighbor on Mammography image,” 2018 18th International Symposium on Communications and Information Technologies (ISCIT),IEEE Xplore.doi:10.1109/ISCIT.2018.8587920
[7.] Tevar Durgadevi Murugan,Mahendra G.Kanojia, ”Breast Cancer Detection Using Texture Features and KNN Algorithm,” In:HIS 2020-Part of the Advances in Intelligent Systems and Computing book series (AISC,volume 1375) ,Springer.doi :10.1007/978-3-030-73050-5-77
[8.] Suhas Athani, Shreesha Joshi, B. Ashwath Rao, Shwetha Rai , N. Gopalakrishna Kini,”Parallel Implementation of kNN Algorithm for Breast Cancer Detection,”Advances in Intelligent Systems and Computing, vol 1176 ,Springer.doi :10.1007/978-981-15-5788-0-46
[9.] Youssef Aamer, Yahya Benkaouz, Mohammed Ouzzif, Khalid Bouragba, ”A new approach for increasing K-nearest neighbors performance,” 2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM),IEEE
[10.] Xplore.doi:10.1109/WINCOM50532.2020.9272459
[11.] G. D. Rashmi, A. Lekha, Neelam Bawane, ”Analysis of Efficiency of Classification and Prediction Algorithms (kNN) for Breast Cancer Dataset,” Advances in Intelligent Systems and Computing, vol 434,Springer.doi:10.1007/978-81322-2752-6-18
[12.] Sonia Goyal , Maheshwar, ”Na¨ıve Bayes Model Based Improved K-Nearest Neighbor Classifier for Breast Cancer
[13.] Prediction,” ICAICR 2019. Communications in Computer and Information Science, vol 1075,Springer.doi:10.1007/978-98115-0108-1-1
[14.] Nur Atiqah Hamzah, Sabariah Saharan , Khuneswari Gopal Pillay ”Classification Tree of Breast Cancer Data with Mode Value for Missing Data Replacement,” Proceedings of the 7th International Conference on the Applications of Science and Mathematics 2021,Springer.doi:10.1007/978-98116-8903-1-25
[15.] Hamim, M., El Moudden, I., Moutachaouik, H., Hain, M, ”Decision Tree Model Based Gene Selection and Classification for Breast Cancer Risk Prediction,” Communications in Computer and Information Science, vol 1207,Springer.doi:10.1007/978-3-030-45183-7-12
[16.] Kriti Jain, Megha Saxena and Shweta Sharma: “Breast Cancer Diagnosis Using Machine Learning Techniques”, IJISET - International Journal of Innovative Science, Engineering Technology, Vol. 5 Issue 5, May 2018.
[17.] Puneet Yadav, Rajat Varshney, Vishan Kumar Gupta: “Diagnosis of Breast Cancer using Decision Tree Models and SVM”. IJISET - International Journal of

Innovative Science, Engineering Technology, Volume: 05 Issue: 03 Mar-2018

[18.] Yixuan Li, Zixuan Chen, 2018: "Performance Evaluation of Machine learning methods for breast cancer prediction", Science publishing group 2018. Doi: 10.11648/j.acm.20180704.15

[19.] Thomas Noel, Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, 2016, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis", Elsevier B.V. 2016.

[20.] Mohammed Amine Naji , Sanaa El Filalib, Kawtar Aarikac, EL Habib Benlahmard, Rachida Ait Abdelouhahide, Olivier Debauchef "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis", Elsevier August 9-12, 2021

[21.] Vishal Deshwal, Mukta Sharma," Breast Cancer Detection using SVM Classifier with Grid Search Technique", International Journal of Computer Applications (0975 – 8887)

[22.] Volume 178 – No. 31, July 2019 Doi: 10.5120/ijca2019919157 [20] S. Sathyavathi, S. Kavitha, R. Priyadharshini and A. Harini, "Breast Cancer Identification Using Logistic Regression" Biosci.Biotech.Res. Comm. Special Issue Vol 13 No 11 (2020) Doi: http://dx.doi.org/10.21786/bbrc/13.11/8

[23.] Prof. Ajit N.Gedam, Kajol B. Deshmane, Nishigandha N.Jadhav, Ritul M.Adhav, Akanksha N.Ghodake," Breast Cancer Detection using Logistic Regression Algorithm", International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Volume 11, Issue 5, May 2022. Doi: 10.15680 /IJIRSET. 2022.1104129

[24.] R. D. Ghongade and D. G. Wakde, "Computer-aided diagnosis system for breast cancer using RF classifier," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 1068-1072 , doi: 10.1109 /WiSPNET. 2017.8299926

[25.] Z. Huang and D. Chen, "A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm," in IEEE Access, vol. 10, pp. 3284-3293, 2022, doi: 10.1109/ACCESS.2021.3139595.

[26.] S. Murugan, B. M. Kumar and S. Amudha, "Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), 2017, pp. 763-766 , doi: 10.1109 /CTCEEC. 2017.8455058

[27.] S. Kabiraj et al., "Breast Cancer Risk Prediction using XGBoost and Random Forest Algorithm," 2020 11 th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-4, DOI: 10.1109 /ICCCNT 49239.2020.9225451.

[28.] R. Fadil, A. Jackson, B. A. El Majd, H. El Ghazi and N. Kaabouch, "Classification of Microcalcifications in Mammograms using 2D Discrete Wavelet Transform and Random Forest," , doi: 10.1109 /EIT 4899

[29.] L. Liu, "Research on Logistic Regression Algorithm of Breast Cancer Diagnose Data by Machine Learning," 2018 International Conference on Robots Intelligent Systems (ICRIS), 2018, pp. 157-160, doi: 10.1109 /ICRIS. 2018.00049.

[30.] B. Dai, R. -C. Chen, S. -Z. Zhu and W. - W. Zhang, "Using Random Forest Algorithm for Breast Cancer Diagnosis," 2018 International Symposium on Computer, Consumer and Control (IS3C), 2018, pp. 449-452, doi:10.1109/IS3C.2018.00119.

[31.] H. Rajaguru and S. K. Prabhakar, "Expectation maximization based logistic regression for breast cancer classification," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 603-606, doi: 10.1109 /ICECA. 2017.8203608.

[32.] Amrita Sanjay , H Vinayak Nair, Sruthy Murali, Krishnaveni K SA Data Mining Model To Predict Breast Cancer Using Improved Feature Selection Method On Real Time Data doi:10.1109/ICACCI.2018.8554450

[33.] Tawseef Ayoub ,Shaikh Rashid Ali Combating Breast Cancer by an Intelligent Ensemble Classifier Approach doi:10.1109/BSB.2018.8770684

[34.] Yogesh Suresh Deshmukh , Parmalik Kumar , Rajneesh Karan , Sandeep K. Singh Breast Cancer Detection-Based Feature Optimization Using Firefly Algorithm and Ensemble Classifier doi:10.1109/ICAIS50930.2021.9395788