Towards Automated Detection of Breast Cancer: Ensemble and Deep Learning Approaches

Md. Saifur Rahman
Dept. of Computer Science and Engineering
East West University
Dhaka, Bangladesh
2018-1-60-048@std.ewubd.edu

Abstract*—Breast cancer affects women a lot by leading them toward death and is the second greatest cause of cancer-related mortality among women. Because of a lack of knowledge and resources, breast cancer is frequently found and diagnosed in its later stages. The most efficient strategy to lower the number of deaths brought on by breast cancer is by early identification and diagnosis. Today, due to the rise in patient volume, manual image processing takes more time and runs the risk of making a false diagnosis. The most trustworthy method for early cancer detection is now machine learning, which may speed up diagnosis and lower the number of fatalities. In this study, we examined breast cancer detection algorithms and chose the most effective one. And the model has been trained and tested using the WDBC. In order to determine which method performs better, we have also side-by-side examined the performance of several algorithms. When we evaluated the effectiveness of Baseline Algorithms, we discovered that Logistic Regression performs the best, with a 98.02% accuracy, 95.24% recall, and 96.71% F1 Score. Linear Discriminant Analysis produces the superior the outcome in terms of precision (99.23%). We have also measured the performance of the Deep Learning algorithms where the CNN gave the best performance in all sectors where the accuracy, precision, recall, and f1 score were 97.89%, 98.30%, 96%, and 97.04% respectively. Finally, when we compared two same types of algorithms, we found that Ensemble approach both hard and soft gave the best performance on accuracy (98.21%), recall (100%), and F1-score (96.55%) but in terms of precision the Random Forest gives the best performance which is 95.47%

***Index Terms*—Breast Cancer, Early detection Logistic Regression, Machine learning, Deep Learning Algorithm, Ensemble approach, and WDBC dataset.**

I. Introduction

One of the most common cancers in women globally and a major contributor to cancer-related death is breast cancer (BC). When breast cells mutate (alter) and overgrow, a mass of tissues is created. Invasive and non-invasive breast cancers are the two main subtypes. Invasive breast cancer refers to breast cancer that has spread to nearby tissues and distant organs. Non- invasive breast cancer does not spread past the milk ducts or breast lobules. Additionally, the appearance of several types of breast cancer under a microscope is used to identify them [1]. Breast cancer claims the lives of numerous people every year. Medical study has identified several risk factors, including age, obesity, hormone imbalance, genetics, a history of radiation exposure, and a family history of breast cancer [2].

A. Motivation

Cancer is a fatal illness that can take a woman's life. Because there is no permanent treatment for cancer and breast cancer is just as harmful as other cancers. According to the WHO report, the number of women who died in 2020 is about 685,000 and predicted that between 2020 and 2040, 2.5 million breast cancer deaths might be avoided if global mortality rates are reduced by 2.5 % per year [3]. However, a number of scientific studies in the medical field have demonstrated that breast cancer may be treated in between 70 and 80 percent of people with early-stage disease, increasing the likelihood of survival. Because early identification of breast cancer enables individuals to obtain effective therapies earlier and lowers the likelihood that they may contract breast cancer [4]. Contrarily, a delayed diagnosis of BC raises the risk of infections from the breast moving to other regions of the body, which increases the risk of contracting cancer and lowers the likelihood of survival because, according to the most recent medical research, cancer is still incurable. Because of unusual symptoms and a doctor's lack of knowledge of breast cancer, it might be challenging to diagnose early. Our study's major goal is to develop a model to better identify breast cancer in its early stages, which will lower mortality. And subsequently, by combining this model with artificial intelligence, we will be able to develop next-generation smart gadgets that can inform a woman about the likelihood of developing breast cancer and what measures may be done sooner to reduce that likelihood.

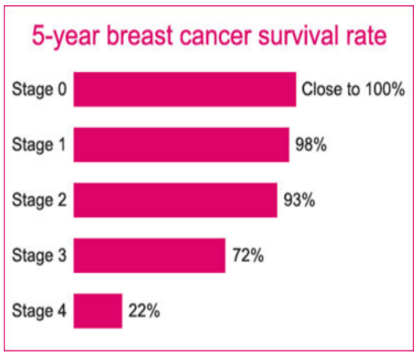

Fig. 1. Survival rate at different stages of Breast cancer [1]

The percentage of people who survive at least 5 years after receiving a breast cancer diagnosis is known as the 5-year breast cancer survival rate.

B. Statement of the Problem

To detect breast cancer, medical science employs a number of techniques, such as CT, ULS, MRI, Histological images of Biopsy, Digital Mammography breast X-ray images (DMG). Digital mammography is frequently used to find the tumor and its boundaries from different perspectives [1].

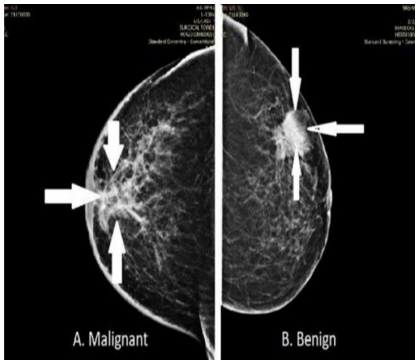

Fig. 2. Survival rate at different stages of Breast cancer [5]

Figure 2 displays two different breast tumor forms (malignant and benign). Using this type of

picture, a radiologist can learn more about the tumor and its condition. The radiologist then thoroughly examines and analyzes them before deciding on the outcomes in consultation with other professionals. This technique takes a long time, and the outcomes heavily depend on the skill and expertise of the radiologist; also, professionals are not readily available everywhere in the world. The last several decades have seen the emergence of machine learning models as a high-accuracy replacement for human vision and judgment in the analysis of medical pictures. Preprocessing, features extraction, and classification are the three main stages of using ML approaches. By deleting any pixels that are unrelated to the tumor, the first stage contributes to an improvement in peripheral visibility and intensity distribution. A critical stage in the identification of breast cancer is feature extraction, which helps to discriminate between benign and malignant tumors. Then, using segmentation, picture properties including regularity, harshness, depth, and smoothness are retrieved. The classification of datasets and the creation of the final product are then completed using various machine learning techniques. Later, a variety of deep learning algorithms are employed to categorize breast cancer. These methods avoid pre-processing and feature extraction, and training and testing on a large dataset before deploying the model for usage can improve the method's performance efficiency [6]. Many baselines and deep learning methods, such as Gradient Boosting, Extreme Gradient Boosting Classifier, Convolutional Neural Network, Recurrent Neural networks, Long Short-Term Memory, Gated Recurrent units, and others, are used to find breast cancer. Additionally, we partitioned the dataset into 10 folds and read it 28 times in order to properly train the model. Later, we calculated the total score's average of accuracy, precision, f1 score, and recall. The ensemble approach was then employed to assess the overall performance of our model.

C. Objective of the study

Building a model that can identify and forecast the likelihood of developing breast cancer from the provided data is the goal of our work. Dataset, which contains information on both benign and malignant tumors, is the source of training and testing data for our proposed model in this work. One of the leading causes of mortality for women and a significant portion of cancer-related deaths is breast cancer. Our study's major goal is to develop a new model that can accurately categorize participants and estimate their likelihood of developing breast cancer. Additionally, minimizing mortality is the goal of this model's creation. Additionally, this model may be combined with AI, which can be used in smart devices to facilitate the earlier identification of breast cancer. However, the current study compares the effectiveness of several methods.

D. Thesis Organization

In section II of our book, we discuss previous breast cancer-related work by different researchers and the basics of machine learning and deep learning. Also, we discussed the basics of the algorithm implemented in our model. The third chapter provides a description of the suggested model, and its subsections go into detail about how it is implemented, including 1.3 Thesis Orientation 3 the visualization of the dataset, data preprocessing, PCA, train-test split, and a brief overview of the algorithms used. The experimental findings and the performance indicators are presented in the next section. The examination and comparison of the outcomes of the various models conclude in chapter four. The last chapter of this essay concludes with a few thoughts about the future.

II. [LITERATURE REVIEW]

Numerous methods, including Naive Bayes, Artificial Neural Network, Support Vector Machine, K Nearest Neighbor, Convolutional Neural Network, and Long Short-Term Memory has been employed by researchers to predict and diagnose

breast cancer. They conducted research on breast cancer utilizing several datasets, including the SEER dataset, the WDBC, WBCD, the MAIS, and the DDSM.

A. Background studies

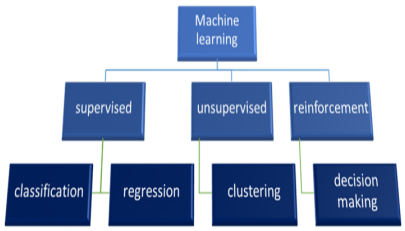

Fig. 3. Classification of Machine learning approach [7]

1) Machine Learning (ML) Approaches: ML is an area of AI that helps computer systems to make more accurate predictions of outcomes. It may be divided into three categories: reinforcement learning, unsupervised learning, and both. Input and output for supervised learning are both labeled so that computers can recognize them. Classification and regression are the main application of supervised learning. But for unsupervised learning, data is not labeled, and the machine learns all the necessary features without any supervision. Unsupervised learning can be used for clustering problems. Reinforcement learning is taking the appropriate action to maximize reward in each situation. It is mostly used to make decisions [7], [8]. Machine learning (ML) has many learning models categories-

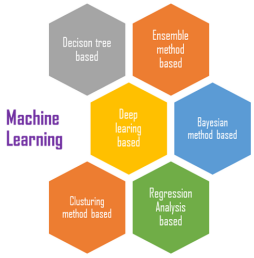

Fig. 4. Machine learning categories [7]

Here, we concentrated on ensemble technique and deep learning in ML. When humans teach machines to do human work so that machines can perform this type of work themselves in the future, it is known as machine learning. But when humans give the machine the ability to mimic the interaction of the human brain is known as Deep learning where the learning pattern of the machine is quite similar to the human brain [9]. A machine learning methodology known as the ensemble method combines several fundamental ML methods [10].

B. Algorithm Details

1) Convolutional Neural Network (CNN): In computer vision, CNN is a subset of deep learning that has gained popularity. It is intended to learn visual feature structures automatically and adaptively utilizing backpropagation and several layers, including fully connected, convolutional, and pooling layers The algorithm has a total of four levels in our implementation, including one input, one output, and several hidden layers for calculation and feature analysis. Although it is normal for some information to be lost during the CNN implementation, padding has been utilized to reduce data loss [11].

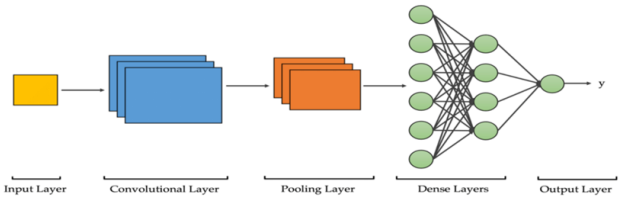

Fig. 5. Structure of CNN [12]

2) Recurrent Neural Network (RNN): RNN can process the sequential data and it can remember its previous learning. So, in our model, the data that has been used for testing can also be used for training in the future state [13].

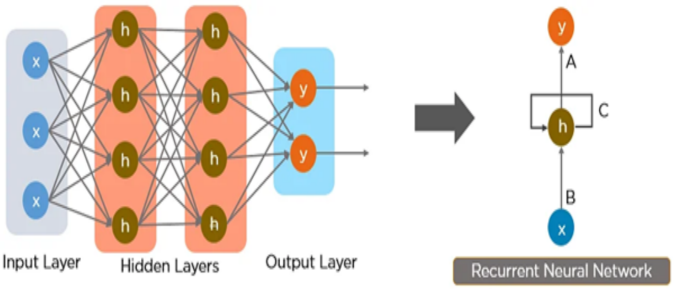

Fig. 6. Concept diagram of RNN [14]

3) Long Short-Term Memory (LSTM): The working format of LSTM is quite like RNN. In our implementation, LSTM learns from its past data and then predicts and this prediction can be saved for future learning this feature has been implemented here [15].

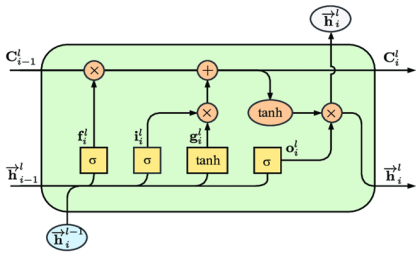

Fig. 7. Concept diagram of LSTM [16]

4) Gated Recurrent Unit (GRU): GRU is the updated version of LSTM, which has internal gates, and that is responsible to control the data flow toward different states. These gates were known as update and reset. The update can determine how much data it will have to recall for optimized learning and a reset can determine how old data needs to be forgotten to avoid overfeeding. So that it can get a well-trained machine to perform its definition perfectly [17].

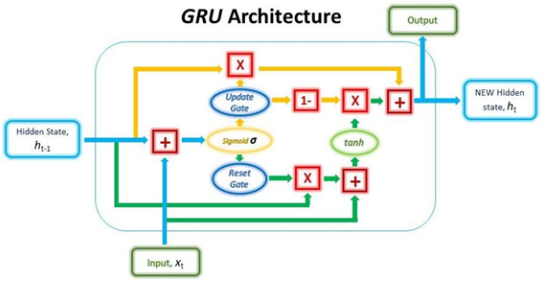

Fig. 8. GRU diagram [18]

5) K Nearest Neighbor (KNN): A non-parametric supervised literacy system is the KNN algorithm. Both retrogression and brackets make considerable use of it. K-NN is a form of division where all figures are deferred until function assessment and the function is only compared locally [19].

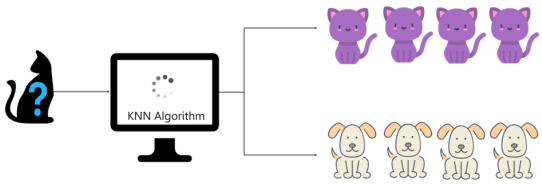

Fig. 9. Example of KNN. [20]

6) Logistic Regression (LR): To forecast the likelihood of a specific goal value, LR is utilized. In general, it may be said that LR is a form of binary classification that determines whether a class is the targeted class defined by '1' or the anticipated class is not the targeted class defined by '0'. But it is divided into multiple categories that can perform different kinds of operations based on the problems i.e., binomial works only for a single class where it indicates either the class is present (1) or the class is not present (0), multinomial is being used only to detect multiple classes but this cannot detect the attribute of the class, ordinal is the update of multinomial where it can detect the class along with the quantitative significance [21].

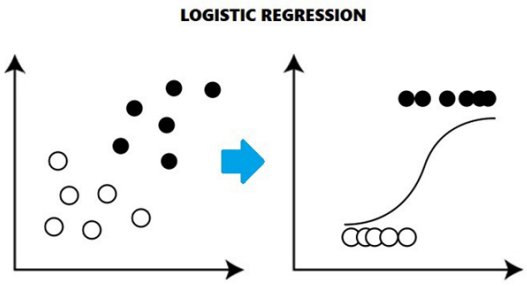

Fig. 10. Logistic Regression Classifier Diagram. [22]

7) Support Vector Classifier (SVC): The SVC approach applies a straightforward kernel function to make division and it performs smoothly with a handsome number of samples. However, the Linear SVC has another parameter like as detriment normalization which applies 'L1' or 'L2' and misplacement function, If we compare it with the SVC model. The kernel system can't be modified in straightforward SVC, because it's predicated on the kernel straight system [23].

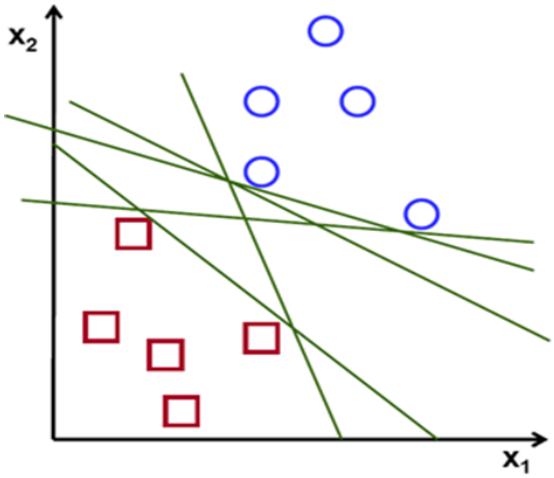

Fig. 11. Support Vector Classifier diagram. [24]

8) Random Forest(RF): A machine learning method called RF combines many categorization methods to address complicated issues. Both problems involving classification and regression may be solved using this method. And this method is employed since it can handle a lot of data while providing the highest level of accuracy. In the same way, precision and recall are comparatively higher than other techniques [25].

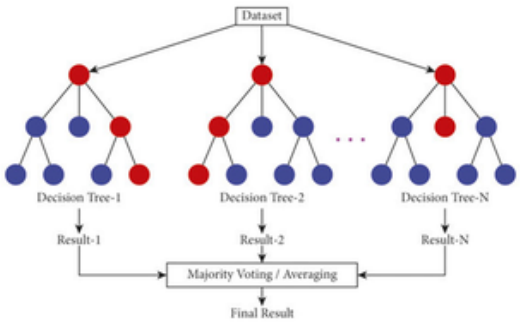

Fig. 12. Random Forest execution process [26]

9) Naïve Bayes: Naïve Bayes Classifier divides the dataset into different categories where it is assumed initially that all the predictions will be independent. This algorithm can perform with less training compared to other algorithms [27].

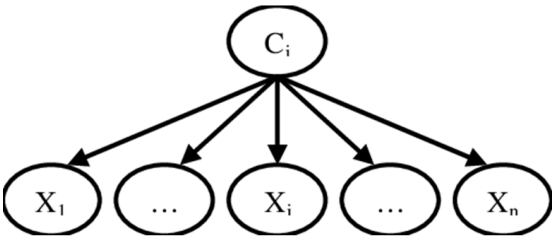

Fig. 13. Structure of Naive Bayes Classifier [28]

10) Ridge Classifier: In situations when linearly independent variables are highly correlated, the Ridge Classifier is

a technique for predicting the coefficients of multiple regression models. It has been utilized extensively in various disciplines, including econometrics, chemistry, and engineering. Ridge regression was developed as a potential solution to the imprecision of the least square an estimator in situations when multicollinear independent variables are included in linear regression models. This provides a more accurate estimate of the ridge parameters since its variance and mean square estimator are typically comparable to the least square estimates previously generated [29].

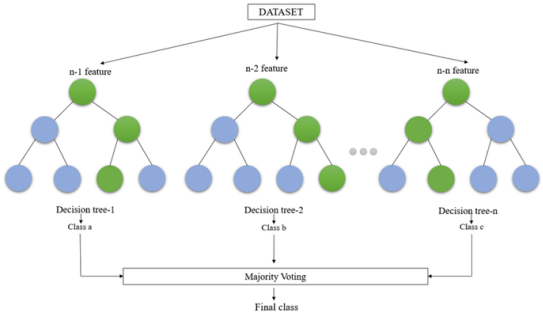

Fig. 14. Structure of Ridge Classifier [30]

11) Gradient Boosting (GB): Among other things, this method is utilized for classification and regression problems. It provides weak ensembles of weak prediction models, often decision trees, in the form of prediction models. GB tree is the name of the resultant method when a decision tree is a weak learner [31].

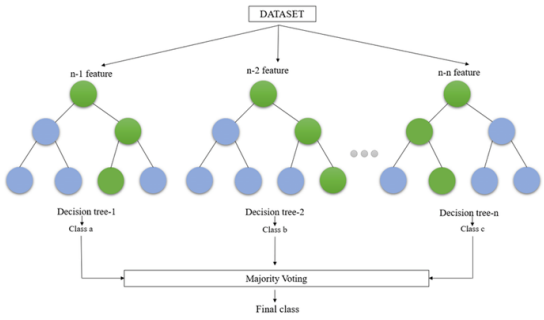

Fig. 15. Gradient Boosted tree [32]

12) Extreme Gradient Boosting Classifier (XGB): Beginning as a terminal program that could be customized via a lib-svm configuration file, XGB Popular features include its sophisticated penalization of trees, proportionate shrinkage of leaf nodes, Newton boosting, additional randomization parameters, implementation of single or networked systems and outdated computation [33].

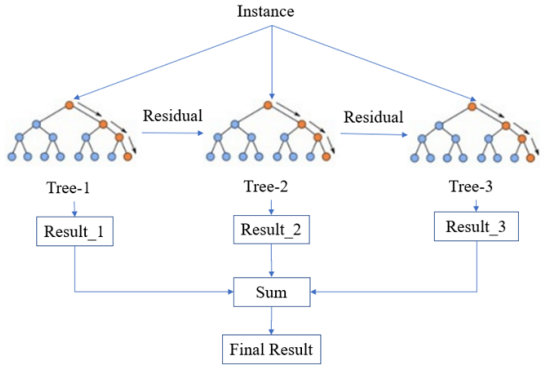

Fig. 16. Structure of XGBoost Classifier [34]

13) Decision Tree: A Decision Tree is a structure that resembles a flowchart and has internal nodes. In a Decision Tree, each internal node stands for a test attribute, and each branch represents the test's outcome. Just as each choice made after computing the characteristics implies a different conclusion, each leaf node reflects a distinct class label. Every path, from the root to the leaves exemplifies categorization concepts. As a visual and analytical decision support tool in decision analysis, the expected values (or expected utility) of computing possibilities are calculated using a decision tree and the closely related impact diagram. Basically, A Decision The tree has three different types of nodes [35]. Basically, there are different types of notations that are used to represent the decision tree.

1) Square is the representation of Decision Node
2) Circle is the representation of Chance Node
3) Triangle represents the End Node

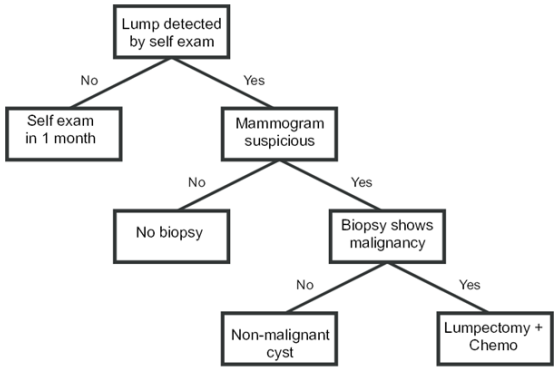

Fig. 17. Example of Decision Tree [36]

14) Linear Discriminant Analysis (LDA): LDA is a technique for reducing the number of dimensions. Before pattern classification, LDA was employed in machine learning as a pre-processing step. LDA aims to reduce dimensionality-related problems and expenses by downscaling features from a higher dimension to lower dimension [37]. Ronald A. Fisher developed an early LDA method known as Fisher's Discriminant Analysis in 1936. It is often referred to as Fisher's Linear Discriminant. A two-class approach was also described. Later, C.R. Rao generalized the Multiple Discriminant Analysis, often known as the Multiple Class LDA. But for the purpose of simplicity, all subsequent techniques are given as LDA [38].

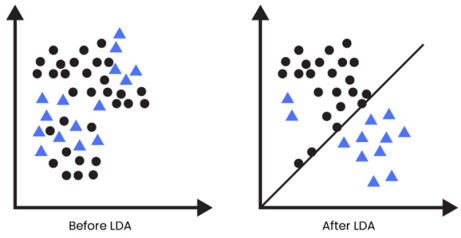

Fig. 18. Linear Discriminant Analysis [39]

C. Ensemble Method

A combination is built up of several independently trained components, such as neural networks or selection trees, whose forecasts are combined to provide a spectacular example. There are two goals for this paper. The most well-known ensemble learning approaches will be reviewed first, along with examples of how they have been used to solve diverse bioinformatics problems. Gene expression, mass spectrometer-based proteomics, discovering gene-gene interactions from genome-wide association studies, and regulatory element prediction from DNA and protein sequences are a few of the primary problems involved [40]. A combination is made up of several independently trained components, such as neural networks or selection trees, whose predictions are combined to provide an elaborate example. Two goals drive this article. First, a summary of the most well-liked ensemble learning approaches will be provided, along with examples of how they have been used to address diverse bioinformatics problems. These problems cover the key fields of gene expression, proteomics using mass spectrometry, detecting gene-gene interactions from genome-wide association studies, and predicting regulatory elements from DNA and protein sequences [41]. The main type of ensemble method:

1) Bagging: Bagging, abbreviated as bootstrap integration, is mainly applied to classification and regression. By using a decision tree to boost the model's accuracy while significantly lowering its diversity. Reduced diversity improves accuracy and gets rid of excess fitting, which is problematic for many prediction models. Bootstrapping and assembly are the two categories under which bagging is categorized. Bootstrapping is a sampling strategy where samples are collected using the replacement approach from the complete population (set) [42].

2) Boosting: Boosting is an integrated approach that improves future prediction by learning from the past. By using a decision tree to boost the model's accuracy while significantly lowering its diversity. Reduced diversity improves accuracy and gets rid of excess fitting, which is problematic for many prediction models. Bootstrapping and assembly are the two categories under which bagging is categorized. Bootstrapping is a sampling strategy where samples are collected using the replacement approach from the complete population (set) [43].

D. Related works

We reviewed Several studies that used machine learning methods for detecting breast cancer. Here are some important research reviews that we followed: Deep Neural Network with Support Value, a novel technique that improves picture quality and addresses other performance criteria, was proposed by [44]. This suggested DNNS technique's main goal is to increase the effectiveness and picture quality for improved prediction and diagnosis. Although in the past, other researchers have also worked with the most popular methods for detecting BC, including the Naïve Bayes, SVM Classifier, Bi-clustering and Ada boost Techniques-CNN Classifier, and HA-BiRNN, and they have had outstanding success in their work. But there are certain drawbacks to these strategies. Therefore, the authors provide a new method or pseudo code called Deep Neural Network with Support Value along with mathematical formulae to measure the effectiveness and performance of DNNS. A DNN, often known as a "deep net" is a neural network with a certain degree of complexity, typically at least two layers. DNN uses advanced math modeling to analyze input in complicated ways. The author employed a sizable dataset that was obtained from the M. G. Cancer Hospital & Research Institute in Visakhapatnam, India. The collection consists of 8009 samples of histopathological images at different magnifications from more than 683 patients. diagnosis. The mathematical formulae used to calculate the histogram value, sigmoid function, and histo- sigmoid functions have been updated and changed for the DNNS approach that has been suggested. Three steps make up the proposed DNNS approach. The input cytology pictures are preprocessed for noise reduction during the pre-processing step. In order to complete this operation, a powerful filtering method was used. The entropy, geometrical, and textural elements are retrieved from the pre-processed pictures in the second step. The breast tumor is separated from the retrieved pictures in the third step. This was accomplished using fuzzy clustering that is Histo-sigmoid based. The operating stance of MATLAB includes the intended DNNS classification. Naive Bayes classifier, SVM classifier, Bi-clustering, and Ada boost techniques, RCNN classifier, Bidirectional Recurrent Neural Networks (HA-BiRNN), and Proposed Methodology - DNN with DNNS all had an average accuracy of 95.61, 95.75, 91.3, 82.50, and 97.21 percent, respectively. The author made sure that the suggested algorithm benefits from performance, efficiency, and picture quality, all of which are significant in today's medical systems. Abien Fred M. Agarap presents a comparison of six ML algorithms: LR, GRU-SVM, MLP, Softmax Regression, SVM, and NN search on the WDBC dataset [45]. Like a lung, bronchus, prostate, clone, and pancreatic cancer, breast cancer is also one of the major cancers all over the world including in the USA. Additionally, applying data science and machine learning can be a wonderful strategy for helping doctors in the medical field. Despite this, other researchers have used the WDBC dataset to study breast cancer in the past and have found great results. This research also addresses the same issue, but it takes a different tack and uses distinct techniques on the same dataset. When training the ML tools, take into account these 10 criteria. - (a) radius, (b) texture, (c) perimeter, (d) area, (e) smoothness, (f) compactness, (g) concavity, (h) concave points, (i) symmetry, and (j) fractal dimension. (mean) (standard deviation) Third-largest mean of the top three values. For its novel suggested technique, this article employed six algorithms, including

1) GRU-SVM: The GRU, the RNN, and the support vector machine are combined in this approach SVM. Parameters have been learned using GRU-RNN and L2-SVM. Additionally, Adam optimization was used to eliminate errors.
2) LR: LR has been used as a classifier for this study

3) MLP: MLP contains multiple hidden layers made with different numbers of a perceptron and ReLU was the activation function for the MLP
4) Nearest Neighbor: This was changed based on the optimization of the problem, and the code implementation was created using Damien's (2017) work from GitHub. Additionally used here for classification are SGD and Adam optimization.
5) Soft-max: This algorithm is used to create the probability distribution for the classes. And the cross-entropy function was used to determine the class
6) SVM: For binary classification—either fall in precise class or not—SVM was utilized. L2-SVM was employed here to get a more stable outcome.

There were two sections to the demonstration: training (70 percent data) and testing (30 percent data). And the requirements were as follows: (1) Test Accuracy (2) Epochs (3) The number of Data Points for assessments. FPR (4.0%) (5) FNR, TPR, TNR, and (6) TNR. Elias Zafiropoulos, Ilias Maglogiannis, and Ioannis Anagnostopoulos conducted all of the trials using computers and applied the SVM with RBF, which had the best test accuracy (89.28The test accuracy for this paper was 96.09 percent, and the test partition was 70/30. GRU- SVM, Linear Regression, MLP, Soft-max, and L2-SVM all had training accuracy averages of 90.6857639%, 92.8906257%, 96.9286785%, 97.366573%, and 97.734375% respectively. Additionally, it took 2:54 minutes, 35, 28, 25 and 14 seconds to train the aforementioned algorithms. The GRU-SVM model was one of several ML algorithms used in this work to calculate the likelihood of breast cancer. And based on the outcome, it was evident that the model was operating with remarkable precision. However, the researchers have suggested using a CV approach, namely k-fold cross validation, to confirm the correctness.

Using gene expression data, Jiande Wu and Chindo Hicks suggested a machine learning technique for separating individuals with triple-negative breast cancer from those without the mutation [46]. When utilizing gene expression data to categorize breast cancer into TNBC and non-TNBC, they tested the effectiveness of four ML-based classification algorithms: KNN, NGB, Decision Tree, and SVM. They asserted that SVM algorithms can properly classify BC into TNBC and non-TNBC. It's crucial to determine if each given woman's body contains the RNA sequence for TNBC or non-TNBC. We need to determine the precise genome sequences of TNBC and non-TNBC in order to answer that question. For such, ML-based systems can outperform more established techniques like ultrasonography and tumor tissue analysis. Additionally, the TCGA dataset was utilized to train an ML system. The RNA genome sequence is required in this dataset, which is publicly accessible, to train the machine to distinguish between TNBC and non-TNBC. This dataset contained 1222 samples and had 60,458 probes. QC and noise reduction have been used at first on initial dataset to eliminate the rows containing inadequate or misplaced data. LIMMA was used on QCed dataset for normalization. We have found in the final dataset 57,179 probes 934 samples of tumor who were distributed as 818 non-TNBC and 116 TNBC samples after data processing and quality control. 90% data randomly have been selected for training and the rest 10% have been used for testing from the final dataset that has been obtained above. Following that, several techniques - SVM, KNN, NGB, and Decision Tree - were applied to the dataset. From the techniques, it can be shown that SVM outperforms KNN, Naive Bayes, and Decision Tree in terms of locating positive and negative samples from the training set. Additionally, the effectiveness of the strategies has been assessed using CM. Cross-validation evaluation has also been performed more times than usual (10 folds) to train the machine better within its limited data. On four distinct plots, the following three metrics - precision, comparison, and category of misclassification—have been computed side by side and used to compare how well, the machine learning algorithms performed:

1) True positive (TP): ML correctly identified the sample as TNBC.
2) True Negative (TN): ML correctly predicted that the sample was non-TNBC.
3) (c)False Positive (FP): The sample was not TNBC despite the ML's prediction that it was TNBC.
4) False Negative (FN): The sample was TNBC notwithstanding the ML's prediction that it wasn't TNBC.

All the above methods were compared with other selection methods like SVM-REF, ARCO, Relief and mRMR. Side by side for application and categorization models to get better results from ML algorithms and performance measurements WEKA has been implemented. After implementing k-NN, NGB, Decision Tree, and SVM, it has been found that in terms of classifying the TNBC and non-TNBC, SVM gave the best result with the highest result on Accuracy, Recall and Specification respectively 90%, 87% and 90%. And NB stayed at the bottom in terms of performance (Accuracy-85%, Recall-68%, Specification-87%). From the paper, we have learned that by using Machine Learning to classify the BC genome and identify the two initial types of genes researchers have got better results. And they have concluded that further research is needed to get advanced machines for identifying these genomes more effectively. Srwa Hasan Abdullah investigated some common ML and DL methods in his essay. They found that SVM's accuracy outperformed other algorithms including K-NN, C4.5, Naïve Bayes, K-means, EM, PAM, and fuzzy c-means by 97 percent [6]. Additionally, the SVM strategy improves accuracy when combined with other techniques like Random Forest, Naive Base, and K-NN. Deeper education has enabled higher accuracy and 98 percent. The three components of Habib Dhahri's work are separated. First, they compared different approaches for selecting features by using machine learning algorithms like PSO and GA [47]. The findings of this experiment showed that 80% of the features were identical. Comparing the performance of machine learning algorithms, choose three methods from KNN, SVM, decision tree, random

forest, AB, GB, GNB, LDA, quadratic discriminant analysis, LR, and extra trees classifier: GB classifier, Random Forest classifier, and extra tree classifier. Finally, they discovered how to use machine learning to automatically create a classifier. More cases of breast cancer detection and diagnosis utilizing different machine learning and deep learning algorithms have been reported.

III. Proposed Model

A. Dataset

Globally In order to train the algorithms for spotting breast cancer, many datasets are employed. Additionally, the WDBC test is frequently employed. Dr. William H. Wolberg, a physician at the University of Wisconsin Hospital in Wisconsin, USA, created this dataset [48]. The dataset was created by Dr. Wolberg using fluid samples obtained from patients with solid breast masses using a straightforward visual computer program called Xcyt to identify cytological characteristics based on a digital scan. This dataset includes the data needed for processing and modeling, including patient IDs and diagnoses (M = malignant, B = benign). Globally In order to train the algorithms for spotting breast cancer, many datasets are employed. Radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension are the ten real-valued properties that are calculated for each of the three cell nuclei. 32 patient specifics, including a note of the patient's ID number, 30 details of the tumor's diagnosis, and 1 note of the tumor diagnosis outcome (benign and malignant).

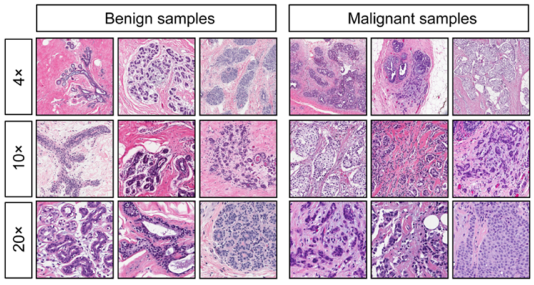

Fig. 19. Histopathology image samples in the WDBC dataset [49]

Attribute Number	Attribute Description	Mean	Standard error	Largest value
1	Radius	6.98-28.11	0.11-2.87	7.93-36.04
2	Texture	0.71-39.28	0.36-4.89	12.02-49.54
3	Perimeter	43.79-188.50	0.76-21.98	50.41-251.20
4	Area	143.50-2501.00	6.80-542.20	185.20-4254.00
5	Smoothness	0.05-0.16	0.00-0.03	0.07-0.22
6	Compactness	0.02-0.35	0.00-0.14	0.03-1.06
7	Concavity	0.00-0.43	0.00-0.40	0.00-1.25
8	Concave points	0.00-0.20	0.00-0.05	0.00-0.29
9	Symmetry	0.11-0.30	0.00-0.08	0.16-0.66
10	Fractal dimension	0.05-0.10	0.00-0.03	0.06-0.21

Fig. 20. WDBC dataset attribute details [50]

B. Implemented Algorithms

We have used a variety of algorithms. Additionally, all algorithms have been tested ten times thanks to the usage of cross-validation that is repeated ten times. We have recorded the results of the iterations'; performance measurements (accuracy, precision, recall, and f1 score). The ten scores we obtained from the 10-fold cross-validation were then averaged.

1) Convolutional Neural Network: We have added all together four layers where the first two layers were the Conv1D() layers with 64 filters and the kernel size was 2 and the activation function was 'Relu'. In the third layer, we have used the Flatten() layer for converting the multi-dimensional matrix into a one- dimensional array. To compile the layers, we used the Adam optimizer and the activation function was sigmoid. Finally, we have trained and tested the model. Then we took the average of all the values that we got from 10-fold cross-validation to determine the final performance.

2) Recurrent Neural Network: In all, three dense layers with sigmoid activation functions were employed in this approach. And as the model was being put together, the binary cross entropy function was employed as the loss measurement function by the Adam optimizer. For training, this model read the dataset a total of 28 times. We evaluated the model's performance by averaging all 10 iterations after ten times of validation.

3) Gated Recurrent Unit: In this implementation, we have used a total of four layers of neural networks. Where every layer takes 50 samples for training. To compile the model, we have used the Adam optimizer, side by side we have also used the accuracy matrix. We have used the mean squared error function to measure the loss. The ten-fold cross-validation to measure the other parameters.

4) Long Short-Term Memory: We have used a total of four layers in the network and every layer takes 50 samples as input to train the model we have used the 'Adam' optimizer, the mean square error was the loss measurement function. Then we have compiled and fit the model for training. Finally, we have tested the models. And, this process has been continued ten times. Then we measured the performance by calculating the average of all the values got from the 10-fold cross-validation.

5) K Nearest Neighbor: We utilized the usual scalar to scale the data when we were using KNN. Bagging Classifiers have also been utilized using KNN. Grid Search CV was used to fine-tune the settings. Base estimator matrix, base estimator weights, and base estimator n-neighbors were the parameters that we employed. Later, we used 10-fold cross-validation to evaluate the model's performance a number of times after training it, with the model reading the dataset 28 times for each validation. We take the average of all numbers following the repetition.

6) Support Vector Classifier: In this implementation, we have used the standard scalar for scaling the data. And we have used the kernel and the gamma parameters for different iterations. We have also used GridSearchCV for hyper tuning the parameters; we have used the 'accuracy' as a scoring parameter. finally, we have launched the iteration 10-fold times. Then we will take the average of all values as the result.

7) Random Forest: In this implementation we have used the parameters, estimators= [10,100] and max_features = ['sqrt','log2'] for different iterations and GridSearchCV for hyperparameter tuning. Then we trained and tested the model. where the testing and training continued for ten iterations. Finally, we have used the average of all parameters for measuring the performance.

8) Ridge Classifier: The parameter we utilized in this implementation is alpha. Then, GridSearchCV was used to tune the hyperparameters. We have utilized 10-fold cross-validation for enhanced accuracy. As an outcome of our training and testing of the model, we utilized the average of all values collected from the iteration.

9) Gradient Boosting: Three parameters were utilized in this implementation: learning rate = [0.001, 0.1, 1], loss = ['deviance', 'exponential'], and n estimators = [10,100,200]. To fine-tune the parameters for the hype, we utilized GridSearchCV. Finally, after repeating this procedure 10 times, the stem fold cross validation was achieved. As an outcome of our training and testing of the model, we utilized the average of all values collected from the iteration.

10) Exclusive Gradient Boosting: Because XGB is already more optimized, we did not use the GridSearchCV in this version to tune the hyperparameters. After that, the k-fold cross validation was performed. Then, to arrive at the ultimate judgment, we must take the average of all values.

11) Logistic Regression: In terms of LR we have used two types of penalties- l1 and l2- then we have also used the GridSearchCV for hyper tuning the parameters. Then we trained and tested the model. Finally, we have repeated the process for 10-fold cross-validation. After that, we measured the average of all values to measure the final assessment.

12) Decision Tree: In terms of the implementation, no hyperparameter has not been used and no such specific parameter has been modified except for some basic function. And the training has been continued 10-fold times. and finally, we will measure the average of all values got from the iteration to give the final assessment

13) Linear Discriminant Analysis: In terms of the implementation, no hyperparameter has not been used and no such specific parameter has been modified except for some basic function. And the training has been continued 10-fold times. and finally, we will measure the average of all values got from the iteration to give the final assessment

14) Quadratic Discriminant Analysis: In terms of the implementation, no hyperparameter has not been used and no such specific parameter has been modified except for some basic function. And the training has been continued 10-fold times. and finally, we will measure the average of all values got from the iteration to give the final assessment

15) Naïve Bayes: In terms of the implementation, no hyperparameter has not been used and no such specific parameter has been modified except for some basic function. And the training has been continued 10-fold times. and finally, we will measure the average of all values got from the iteration to give the final assessment

16) Ensemble Method: Previously we have run all the Bayesian models for 10-fold time to measure their performance but at that time All the assessments were done on those algorithms individually. So, to measure the overall average performance we have used the ensemble voting method where we have used both hard and soft voting methods and have got the average score of all models that we have used previously in the 10-fold cross-validation. By re-evaluating those scores, we have got the performance of the ensemble voting method. Here both hard and soft voting methods have been used. As all the models have been trained earlier with 10-fold cross-validation, no hyper-parameter tuning has been done here.

C. Methodology

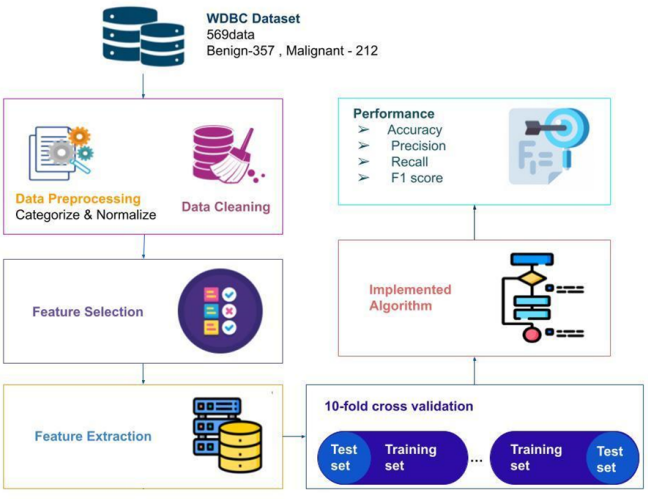

Fig. 21. Workflow for the proposed model.

To train the model, we will have to read the dataset from the WDBC dataset CSV file. Then we will drop unnecessary columns i.e., id and unnamed 32 as those are not necessary for training the model so that these columns cannot harm the

accuracy of the model. But the presence of these columns takes extra memory and processing power. After that, we will reshape the dataset by removing the null values, because these null values create anomalies that will be responsible for bad training and so to train the model better and get the better performance it is necessary to remove all the null values. Outliers are another problem for training an ML model. When there are a lot of outliers in the dataset that time machine- while it was training itself- cannot hyper-tune its parameters to prepare itself for the work. And while the machine will be deployed for the operation, it will not be able to perform its work properly. So, dealing with the outliers is necessary for better training. And, we will handle the outliers by scaling them down. Later the main work remains- extracting the features from the dataset. Without knowing the features, we will not be able to train our model. So, we will extract the features by defining 'inputs' and define the target as 'targets. The dataset will next need to be divided into training and testing portions for the model. Because we only have one data set and we employed 10- fold cross-validation, our dataset was divided into 10 pieces for each iteration, with 9 portions being used for training and 1 section for training. Then we will train the model using different algorithms in the loop where the training and testing process will be continued 10 times using the datasets, where the value of epochs is 60. Then the results of all iterations will be appended into a list for further calculations. The average of all the findings from the k-fold cross-validation will then be determined. such that the accuracy, precision, recall, and F1-score values are more accurate. We will evaluate the overall effectiveness of accuracy, precision, recall, and F1 score for confirmation using the Ensemble approach.

IV. Result Analysis

A. Performance Metrics

1) Confusion Matrix: The number of properly-recognized tuples and the number of incorrectly identified tuples may be found in a table called a confusion matrix, which displays the performance of an algorithm [51]. There are four methods to partition the confusion identification type matrix. Terms and Conditions about Confusion Matrix:

1) TP: TP refers to the number of samples that the algorithm or ML model correctly recognized as positive and those samples truly were positive [52].
2) TN: TN means the number of samples identified as negative by the algorithm/ ML model and the samples were actually negative [52].
3) FP: FP are samples that the algorithm or ML model incorrectly recognized as positive when they were actually negative samples. It's also referred to as a 'Type I error' [52].
4) FN: FN means the number of samples identified as negative by the algorithm/ ML model and the samples was not actually negative. It is also known as a 'Type II error' [52]

Prediction outcome

actual value	**p**	**n**	**total**
p′	True Positive	False Negative	P′
n′	False Positive	True Negative	N′
total	P	N	

2) Accuracy: In general, accuracy refers to how well the model performs across all classes [53]. When all classes are equally important, it is helpful. The number of accurate forecasts divided by the total number of predictions is used to compute it. The formula for accuracy is [54] :

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

3) Precision: The precision is determined by dividing the total number of Positive samples (classified either properly or wrongly) by the number of Positive samples that were correctly categorized [55]. The precision gauges how well the model categorizes a sample as positive. The individuals with malignant tumors who were actually found to have them were the expected positives (TP and FP). Precision is defined as follows [56]:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

4) Recall or Sensitivity: The recall serves as a barometer for how well our approach detects TP [57]. Recall, therefore, informs us how many of the patients we can theoretically identify as having illness out of all the ones who really do. Therefore, we would want our Recall to be as near to 100 percent as feasible if we were to concentrate more on reducing False Negatives. The recall is defined as follows [56] :

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

5) F1-Score: The accuracy of a model on a dataset is gauged by the F-score, also known as the F1-score. It is employed to assess binary classification systems that categorize examples as either positive or negative [58]. The linear mean of the model's accuracy and recall is known as the F-score, which combines these two metrics [59]:

$$F1\_Score = \frac{\text{Precision} \times 2}{\text{Precision} + \text{Recall}}$$

B. Model Performance

1) Support Vector Classifier: After applying 10-fold cross-validation the accuracy we have found from SVC is 97.18%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the

correctness of forecasts for underrepresented groups. Here precision is 97.85%. And the Recall and F1 scores are 94.80% and 96.20% respectively.

2) Random Forest: After applying 10-fold cross-validation the accuracy we have found from RF is 96.30%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 95.47%. And the Recall and F1-score are 94.67% and 94.92%.

3) Ridge Classifier: After applying 10-fold cross-validation the accuracy we have found from RC is 95.95%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 99.23%. And the Recall and F1 scores are 89.68% and 94.17%.

4) Gradient Boosting: After applying 10-fold cross-validation the accuracy we have found from GB is 97.01%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 98.59%. And the Recall and F1-score are 93.67% and 96.20% respectively.

5) Exclusive Gradient Boosting : After applying 10-fold cross-validation the accuracy we have found from XGB is 97.18%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision, Recall, and F1-score are 97.47%, 95.13%, and 96.15%.

6) Logistic Regression: After applying 10-fold cross-validation the accuracy we have found from LR is 97.54%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 98.38%. And the Recall and F1 scores are 95.24% and 96.71% respectively.

7) Linear Discriminant Analysis: After applying 10-fold cross-validation the accuracy we have found from LDA is 95.60%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 99.23%. And the Recall and F1-score are 88.78% and 93.65% respectively

8) Quadratic Discriminant Analysis: After applying 10-fold cross-validation the accuracy we have found from QDA is 95.60%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 93.59%. And the Recall and F1-score are 94.70% and 94%.

9) K Nearest Neighbor: After applying 10-fold cross validation the accuracy we have found from KNN is 96.48%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 97.19%. And the Recall and F1-score are 93.28% and 95.68% respectively.

10) Naïve Bayes: After applying 10-fold cross validation the accuracy we have found from Naïve Bayes is 93.49%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 91.96%. And the Recall and F1-scores are 90.52% and 91% respectively.

11) Decision Tree: After applying 10-fold cross validation the accuracy we have found from the DT is 91.38%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 87.06%. And the Recall and F1-score are 89.69% and 88.01%.

12) Convolutional Neural Network: After applying 10-fold cross-validation the accuracy we have found from RF is 97.89%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 98.30%. And the recall and F1-score are 95.99% and 97.04% respectively.

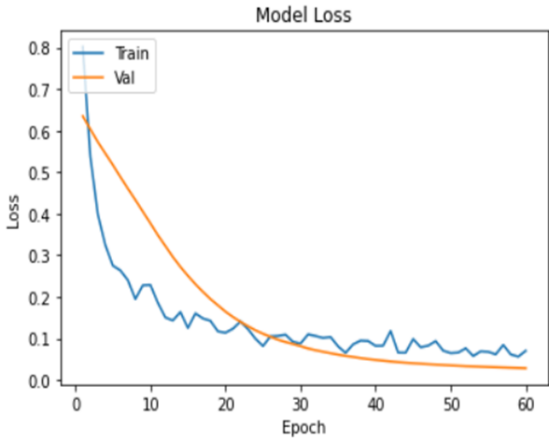

Fig. 22. Loss function graph of CNN.

13) Recurrent Neural Network: After applying 10-fold cross-validation the accuracy we have found from RNN is 96.65%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 95.88%. And the Recall and F1-score are respectively 95.02% and 95.24%

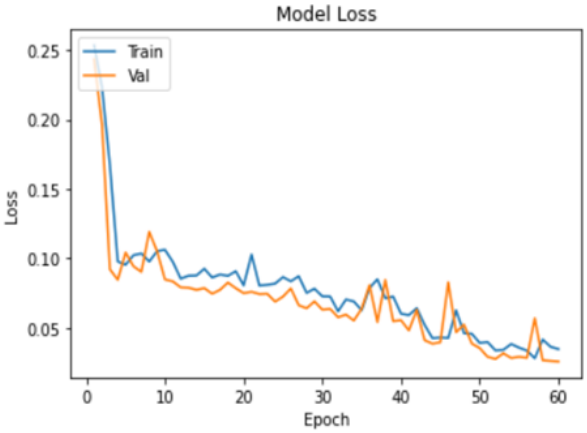

Fig. 23. Loss function graph of RNN.

14) Gated Recurrent Unit: After applying 10-fold cross-validation the accuracy we have found from GRU is 95.77%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision score is 92.92% with recall and F1-Score are respectively 95.13% and 93.91%.

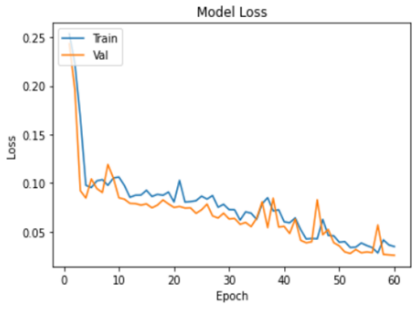

Fig. 24. Loss function graph of GRU.

15) Long Short-Term Memory: After applying 10-fold cross-validation the accuracy we have found from LSTM is 94.54%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision is 93.13%. And the Recall and F1-score are respectively 91.81% and 92.19%

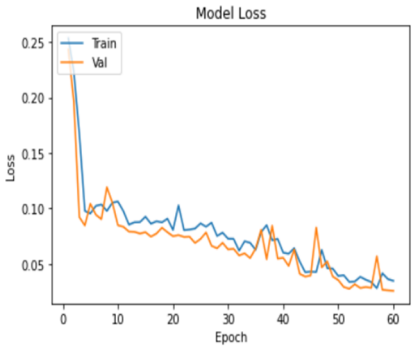

Fig. 25. Loss function graph of LSTM.

16) Ensembling Voting Method: This method has been divided into two sections:

1) Ensemble Soft Voting: After applying 10-fold cross-validation the accuracy we have found from Ensemble Soft is 98.21%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision score is 93.33% with recall and F1-Score are respectively 100% and 96.55%.
2) Ensemble Hard Voting: After applying 10-fold cross-validation the accuracy we have found from Ensemble Hard is 98.21%. We know that Precision is inversely correlated with the accuracy of all correctly predicted positive outcomes, or the correctness of forecasts for underrepresented groups. Here precision score is 93.33% with recall and F1-Score are respectively 100% and 96.55%.

If we compare the overall performance among the baseline algorithms, the LR gives the best result on accuracy (97.54%), Recall (96%) and F1 Score. In terms of Precision, RC and LDA give the best results (99.23%). In the same way, if we measure the performance of F1 score(96.71%) but for precision the LDA give the the highest result (99.23%)

If we compare the overall performance, the CNN gives the best result on accuracy (97.89%), Precision (98.30%), Recall (96%), and F1 score (97.04%). In the second position of overall performance, RNN gives the result on accuracy (96.65%), Precision (95.88%) and F1 score (95.24%). And GRU occupies the third position

We discovered that the Ensemble strategy, which uses both hard and soft voting methods, produces the best results overall

TABLE I
BASELINE MACHINE LEARNING ALGORITHMS' RESULT

Model	Accuracy	Precision	Recall	F1 Score
Ridge Classifier	95.95%	99.23%	89.68%	94.17%
Gradient Boosting Classifier	97.01%	98.58%	93.67%	96.02%
XGBoost	97.18%	97.49%	95.13%	96.15%
Logistic Regression	97.54%	98.38%	95.24%	96.71%
Decision Tree Classifier	91.38%	87.06%	89.69%	88.01%
Linear Discriminant Analysis	95.60%	99.23%	88.78%	93.65%
Quadratic Discriminant Analysis	95.60%	93.59%	94.70%	94%
Naive Bayes Classifier	93.49%	91.96%	90.52%	91%
Support vector classifier	97.18%	97.85%	94.80%	96.20%
K-nearest neighbors	96.48%	97.19%	93.28%	95.06%

TABLE II
DEEP LEARNING ALGORITHMS' RESULT

Model	Accuracy	Precision	Recall	F1 Score
Convolutional Neural Network	97.89%	98.30%	96%	97.04%
Recurrent Neural Network	96.65%	95.88%	95.02%	95.24%
Long Short-Term Memory	94.54%	93.13%	91.81%	92.19%
Gated Recurrent Unit	95.77%	92.92%	95.13%	93.91%

in terms of accuracy, recall, and F1 score, with scores of 98.21 percent, 100 percent, and 96.55 percent, respectively. However, Random Forest performs best when measuring performance on the Precision side, with a Precision result of 95.47 percent.

V. CONCLUSION & FUTURE WORKS

DL and ML techniques can support early cancer detection. The models for cancer detection employ a wide variety of algorithms. We selected these algorithms as the best based on accuracy, precision, recall, and f1 score. This study demonstrated a comparison of 15 algorithms for MRI breast cancer detection under Bayesian and DL methods. We see that under the Baseline and deep learning algorithms, the GBC and RNN perform better in terms of accuracy, recall, and F1-Score. In the end, we determined that the GBC offers the greatest performance with an accuracy rate of 98.25 percent. An expert doctor can use this approach to quickly and accurately reach a judgment. With more data being available in the future, our model may be improved, and we work to raise the accuracy level. Because we have failed to incorporate the advanced feature into our model which is why we do not get the best results from our model. Though some parts of our model do not perform well, we still hope that if we can incorporate more advanced features like refined hyperparameter tuning, our model will be able to perform better. We will incorporate transfer learning and other hybrid algorithms into this model for improving the model and its performance [60]. And our improved model will also be implemented on different datasets as well for further improvement and better training of our model. This model can be incorporated with the Al and can be used in the smart machines, which will help the machines to detect the probability of being infected by breast cancer with more accuracy. Due to the limitations of our machine, we were not able to increase the epoch by more than 60. As the processing power was not sufficient, our machine cannot compute the bigger calculation with a large epoch. So we will try to increase the epoch number.

TABLE III
ENSEMBLE METHODS COMPARISON RESULT

Model	Accuracy	Precision	Recall	F1 Score
Ensemble Soft	98.21%	93.33%	100%	96.55%
Ensemble Hard	98.21%	93.33%	100%	96.55%
Random Forest	96.30%	95.47%	94.67%	94.92%

REFERENCES

[1] "Breastcancer.org." https://www.breastcancer.org/. [Online; accessed 2022-07-22].

[2] Y.-S. Sun, Z. Zhao, Z.-N. Yang, F. Xu, H.-J. Lu, Z.-Y. Zhu, W. Shi, J. Jiang, P.-P. Yao, and H.-P. Zhu, "Risk Factors and Preventions of Breast Cancer," *International Journal of Biological Sciences*, vol. 13, no. 11, pp. 1387–1397, 2017. [Online; accessed 2022-07-22].

[3] W. H. O. WHO, "Breast cancer," *World Health Organization: WHO*, mar 26 2021. [Online; accessed 2022-07-22].

[4] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, and F. Cardoso, "Breast cancer is a systemic disease optimally treated by a multidisciplinary team," *Nature Reviews Disease Primers*, vol. 5, sep 23 2019. [Online; accessed 2022-07-22].

[5] T. Mahmood, J. Li, Y. Pei, F. Akhtar, A. Imran, and K. U. Rehman, "A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities," *IEEE Access*, vol. 8, pp. 165779–165809, 2020.

[6] S. H. Abdulla, A. M. Sagheer, and H. Veisi, "Breast cancer classification using machine learning techniques: A review," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 14, pp. 1970–1979, 2021.

[7] Z. Ghahramani, *Unsupervised Learning*, pp. 72–112. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. [Online; accessed 2022-07-22].

[8] S. Gollapudi, *Practical machine learning*. Packt Publishing Ltd, 2016.

[9] L. Deng and D. Yu, "Foundations and trends in signal processing: Deep learning–methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3-4, pp. 197–387, 2014. [Online; accessed 2022-07-22].

[10] L. Rokach, *Ensemble Methods in Supervised Learning*, pp. 959–979. Rokach, Lior: Springer US, 2009. [Online; accessed 2022-07-22].

[11] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 24–49, 3 2021. [Online; accessed 2022-07-22].

[12] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, pp. 611–629, jun 22 2018. [Online; accessed 2022-07-22].

[13] S. Agrawal and J. Agrawal, "Neural Network Techniques for Cancer Prediction: A Survey," *Procedia Computer Science*, vol. 60, pp. 769–774, 2015. [Online; accessed 2022-07-22].

[14] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 3 2020. [Online; accessed 2022-07-22].

[15] R. K. Bhogal, P. D. Suchit, and C. Naresh, "Review: Breast Cancer Detection Using Deep Learning," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, jun 3 2021. [Online; accessed 2022-07-22].

[16] A. Das and M. N. Mohanty, "A review of breast cancer detection using deep learning techniques," *Advances in Deep Learning for Medical Image Analysis*, pp. 69–85.

[17] M. H. Kamarudin, Z. H. Ismail, and N. B. Saidi, "Deep Learning Sensor Fusion in Plant Water Stress Assessment: A Comprehensive Review," *Applied Sciences*, vol. 11, p. 1403, feb 4 2021. [Online; accessed 2022-07-22].

[18] A. E. Carrington, N. Solomon, and R. K. Sivamani, "Squamous Cell Carcinoma of Unknown Primary," *Consultant*, 2021. [Online; accessed 2022-07-22].

[19] A. Bharat, N. Pooja, and R. A. Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis," in *2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C)*, IEEE, 10 2018. [Online; accessed 2022-07-22].

[20] J. C. Konje, *Examples of Possible Tasks under the Various Domains*, pp. 15–16. CRC Press, nov 15 2020. [Online; accessed 2022-07-22].

[21] M. Nourelahi, A. Zamani, A. Talei, and S. Tahmasebi, "A model to predict breast cancer survivability using logistic regression," *Middle East Journal of Cancer*, vol. 10, no. 2, pp. 132–138, 2019.

[22] R. Vidhya, P. Karthik, and S. Jamadagni, "Anticipatory QoE Mechanisms for 5g Data Analytics," in *2020 International Conference on COMmunication Systems amp; NETworkS (COMSNETS)*, IEEE, 1 2020. [Online; accessed 2022-07-22].

[23] V. Chaurasia and S. Pal, "Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer," *SN Computer Science*, vol. 1, pp. 1–11, aug 14 2020. [Online; accessed 2022-07-22].

[24] Leu-Shing Lan, "M-SVC (Mixed-Norm SVC) — A Novel Form of Support Vector Classifier," in *2006 IEEE International Symposium on Circuits and Systems*, IEEE. [Online; accessed 2022-07-22].

[25] B. Dai, R.-C. Chen, S.-Z. Zhu, and W.-W. Zhang, "Using Random Forest Algorithm for Breast Cancer Diagnosis," in *2018 International Symposium on Computer, Consumer and Control (IS3C)*, pp. 449–452, IEEE, 12 2018. [Online; accessed 2022-07-22].

[26] M. Y. Khan, A. Qayoom, M. S. Nizami, M. S. Siddiqui, S. Wasi, and S. M. K.-u.-R. Raazi, "Automated prediction of good dictionary examples (gdex): A comprehensive experiment with distant supervision, machine learning, and word embedding-based deep learning techniques," *Complexity*, vol. 2021, 2021.

[27] M. Oujaoura, B. Minaoui, and M. Fakir, "Combined descriptors and classifiers for automatic image annotation," in *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, vol. 8, pp. 225–244, Ieee, Citeseer, 4 2014. [Online; accessed 2022-07-22].

[28] A. A. Bataineh, "A Comparative Analysis of Nonlinear Machine Learning Algorithms for Breast Cancer Detection," *International Journal of Machine Learning and Computing*, vol. 9, pp. 248–254, 6 2019. [Online; accessed 2022-07-22].

[29] R. R. Kadhim and M. Y. Kamil, "Comparison of breast cancer classification models on wisconsin dataset," *Int J Reconfigurable & Embedded Syst ISSN*, vol. 2089, no. 4864, p. 4864.

[30] H. Binte Kibria and A. Matin, "The severity prediction of the binary and multi-class cardiovascular disease–a machine learning-based fusion approach," *arXiv e-prints*, pp. 2200–2203, 2022.

[31] A. Kamboj, P. Tanay, A. Sinha, and P. Kumar, *Breast Cancer Detection Using Supervised Machine Learning: A Comparative Analysis*, pp. 263–269. Singapore: Springer Singapore, sep 17 2020. [Online; accessed 2022-07-22].

[32] M. Shoaran, B. A. Haghi, M. Taghavi, M. Farivar, and A. Emami-Neyestanak, "Energy-Efficient Classification for Resource-Constrained Biomedical Applications," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, pp. 693–707, 12 2018. [Online; accessed 2022-07-22].

[33] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Machine Learning with Applications*, vol. 6, p. 100154, 12 2021. [Online; accessed 2022-07-22].

[34] W. Wang, G. Chakraborty, and B. Chakraborty, "Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm," *Applied Sciences*, vol. 11, p. 202, dec 28 2020. [Online; accessed 2022-07-22].

[35] J. Villalba, N. Abdussamie, V. Aryai, N. Nikolova, K. Tenekedjiev, C.-M. Wang, and I. Penesis, "Assessment of uncertain alternatives for co-located aquaculture and offshore wind farm in tasmania," *Ocean Engineering*, vol. 249, p. 110949, 4 2022. [Online; accessed 2022-07-22].

[36] J. A. Cruz and D. S. Wishart, "Applications of Machine Learning in Cancer Prediction and Prognosis," *Cancer Informatics*, vol. 2, p. 117693510600200, 1 2006. [Online; accessed 2022-07-22].

[37] *Discriminant Analysis (DA) and Other Knowledge-Based Classification Methods*, pp. 269–280. Chichester, UK: John Wiley Sons, Ltd. [Online; accessed 2022-07-22].

[38] A. Alzahrani and D. B. Rawat, "Comparative Study of Machine Learning Algorithms for SMS Spam Detection," in *2019 SoutheastCon*, Ieee, 4 2019. [Online; accessed 2022-07-22].

[39] N. Tyagi, *Linear regression and discriminant analysis for supervised classification*, pp. 248–281. Cambridge University Press, apr 17 2014. [Online; accessed 2022-07-22].

[40] P. M. Mitrasinovic, *Contemporary Trends in Bioinformatics Relevant for Some Important Biomedical Problems*. Bentham Science Publishers, dec 15 2012. [Online; accessed 2022-07-22].

[41] W. Wang, "Some fundamental issues in ensemble methods," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Ieee, 6 2008. [Online; accessed 2022-07-22].

[42] P. Bühlmann, *Bagging, Boosting and Ensemble Methods*, pp. 985–1022. Berlin, Heidelberg: Springer Berlin Heidelberg, dec 21 2011. [Online; accessed 2022-07-22].

[43] O. Sprangers, S. Schelter, and M. de Rijke, "Probabilistic Gradient Boosting Machines for Large-Scale Probabilistic Regression," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery amp; Data Mining*, (New York, NY, USA), pp. 1510–1520, ACM, aug 14 2021. [Online; accessed 2022-07-22].

[44] A. R. Vaka, B. Soni, and S. R. K., "Breast cancer detection by leveraging Machine Learning," *ICT Express*, vol. 6, pp. 320–324, 12 2020. [Online; accessed 2022-07-22].

[45] A. F. M. Agarap, "On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset," in *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing - ICMLSC '18*, (New York, New York, USA), pp. 5–9, ACM Press, 2018. [Online; accessed 2022-07-22].

[46] J. Wu and C. Hicks, "Breast Cancer Type Classification Using Machine Learning," *Journal of Personalized Medicine*, vol. 11, p. 61, jan 20 2021. [Online; accessed 2022-07-22].

[47] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–11, nov 3 2019. [Online; accessed 2022-07-22].

[48] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Breast cancer wisconsin (diagnostic) data set [uci machine learning repository]," 1992.

[49] A. Pon Bharathi, Mr, D. Pallavi, Mrs, M. Ramachandran, Mr, R. Kurinjimalar, Miss, and P. Vidhya, Mrs, "A Study on Preference Selection Index Multi-Criteria Decision Making Techniques," *Data Analytics and Artificial Intelligence*, vol. 2, pp. 20–25, mar 1 2022. [Online; accessed 2022-07-22].

[50] H. Lin and Z. Ji, "Breast Cancer Prediction Based on K-Means and SOM Hybrid Algorithm," *Journal of Physics: Conference Series*, vol. 1624, p. 042012, oct 1 2020. [Online; accessed 2022-07-22].

[51] J. Brownlee, *Machine Learning Algorithms From Scratch with Python*. Machine Learning Mastery, nov 16 2016. [Online; accessed 2022-07-22].

[52] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 6 2006. [Online; accessed 2022-07-22].

[53] P. Eusebi, "Diagnostic Accuracy Measures," *Cerebrovascular Diseases*, vol. 36, no. 4, pp. 267–272, 2013. [Online; accessed 2022-07-22].

[54] M. Sokolova, N. Japkowicz, and S. Szpakowicz, *Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation*, pp. 1015–1021. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. [Online; accessed 2022-07-22].

[55] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, 2020.

[56] D. L. Olson and D. Delen, *Advanced data mining techniques*. Springer Science & Business Media, 2008.

[57] D. Gur, J. H. Sumkin, H. E. Rockette, M. Ganott, C. Hakim, L. Hardesty, W. R. Poller, R. Shah, and L. Wallace, "Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System," *JNCI Journal of the National Cancer Institute*, vol. 96, pp. 185–190, feb 3 2004. [Online; accessed 2022-07-22].

[58] Y. Sasaki *et al.*, "The truth of the f-measure," *Teach tutor mater*, vol. 1, no. 5, pp. 1–5, 2007.
[59] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, jan 2 2020. [Online; accessed 2022-07-22].
[60] A. Mari, T. R. Bromley, J. Izaac, M. Schuld, and N. Killoran, "Transfer learning in hybrid classical-quantum neural networks," *Quantum*, vol. 4, p. 340, 2020.