A Comparative Study of Machine Learning Algorithms Applied to Cancer Diagnosis Specific to Breast Cancer

Ankur Gupta[1*], Rajdeep Singh[2], Krishan Kumar[3], Piyush Kumar[4], Arvind Kumar Shukla[5]

[1,2] School of Computer Science and Applications, IFTM University, Moradabad, U.P., India

[3]Department of Computer Science, Gurukula Kangri (Deemed to be University), Haridwar, 249404, Uttarakhand, India

[4]Department of Radiation Oncology, Shri Ram Murti Smarak Institute of Medical Sciences, Bareilly, U.P., India

[5] Department of CS & IT, Mahatma Gandhi Central University, Motihari, Bihar (India)

Corresponding Author Email: Ankur Gupta

Abstract

Improvement of the fact at which patients' survival and treatment results break is that which we see out of very early detection of breast cancer. Presently we see that traditional diagnostic methods do in to play to that of human subjectivity, tiredness as a factor in performance and also variable interpretation which is a large issue because they mostly use radiology and pathology for evaluation. These issues put into light the need for data based, automated approaches which in turn will give out very reliable and unbiased results. To have an objective performance evaluation we pre-processed the breast mass image data which included features from fine-needle aspirates. We used the same settings for training and testing each model to ensure evaluation consistency. We found that ensemble models which include Random Forest and Gradient Boosting performed the best out of the studied algorithms with over 98.8% accuracy. Also, these models had better recall and AUC which in turn showed their performance in differentiation between benign and malignant cases. Also, we saw that which statistical analysis, confusion matrices, and descriptive visualizations. What we found is that ensemble-based machine learning does in fact out perform traditional models in the case of breast cancer which we put forth as a reliable and scalable solution for early detection. In health care settings this type of predictive modelling may improve the accuracy of cancer screen out comes, reduce in diagnostic errors, and support clinical decision making.

Keywords: *Machine Learning, Feature Importance, ROC Curve, Classification, Breast Cancer, and Comparative Analysis*

I. Introduction

Each year breast cancer reports as a very large cause of death from cancer which also has the highest incidence and is the greatest health issue in the field of oncology [1], [5]. We see that early and accurate diagnosis is what really improves outcomes which in turn see to better results in terms of survival. Although we have seen great technical growth in the field of diagnostics, we still see mainly the use of mammography, ultrasounds, and histopathological analysis which at their root still require the very subjective manual interpretation by radiologists and pathologists which in turn brings in issues of variable diagnosis between readers and also in some cases diagnostic error which may be a result of human fatigue [6]. Also, these issues play a large role in the quality of diagnosis in very busy clinical settings and in health care structures which are limited in resource [7].

In Fig. 1 we present a case of breast cancer which displays the most tell tissue feature of cancer a spiculated mass. What we see is that when a tumor produces fibrotic tissue which extends out like spikes the result is spiculated margins which in turn point to invasive growth. Thus, these imaging features stress the importance of accurate image interpretation.

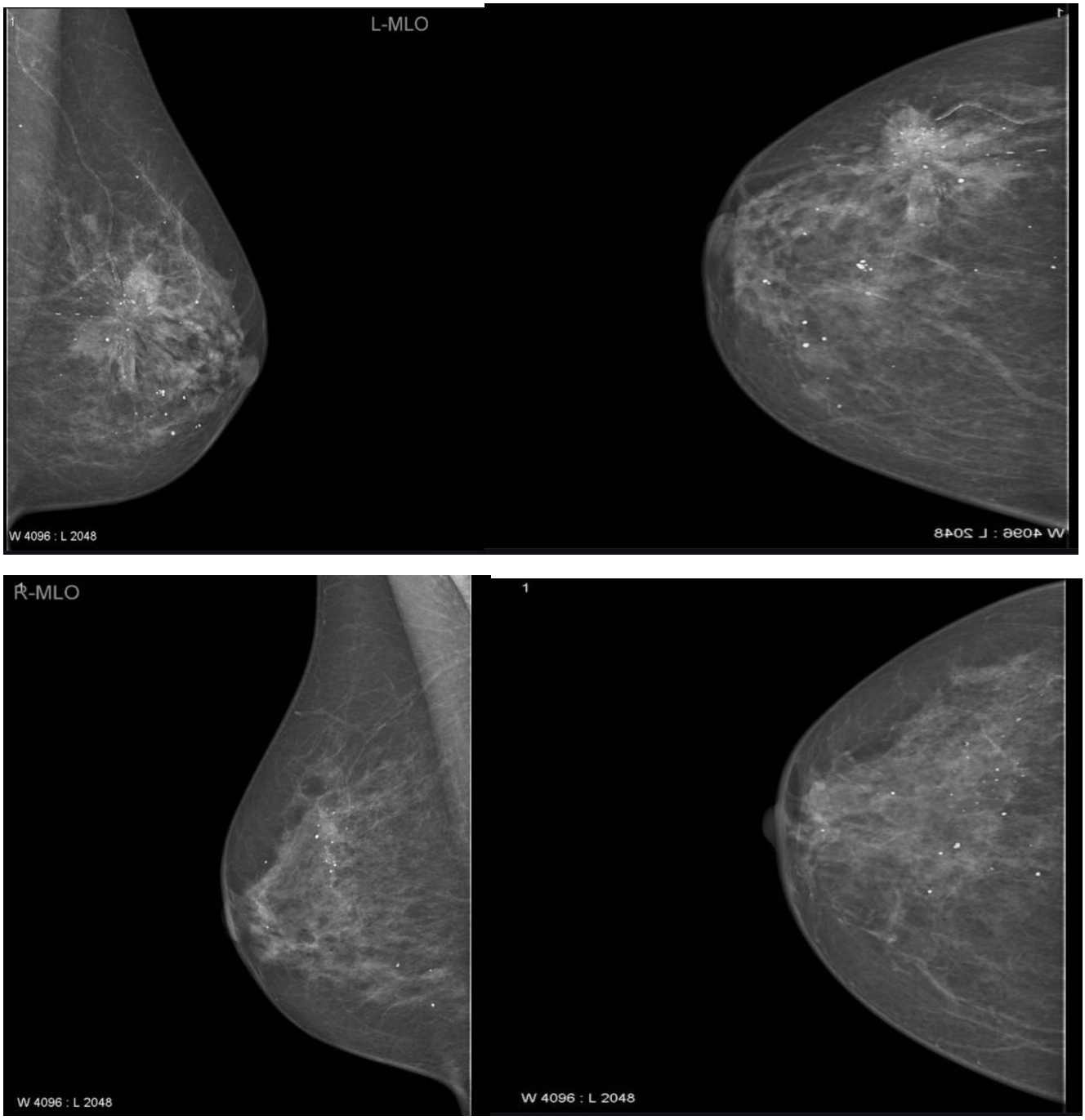


Figure 1: Mammography Image of a Spiculated Breast Tumor Signifying Invasive Carcinoma

The mammogram in Figure 1 displays a high density spiculated tumor which is a sign of invasive breast cancer [36]. In this image we see a mammography case of a spiculated mass that very much is related to malignant breast cancer. What is shown are strong radiating projections which also serve as a clinical reference for diagnosis. In recent years Machine Learning grows very fast that shows accurate results comparison of manual methods. The algorithms of machine learning help doctors because they make accurate predictions by learning from complex and non-linear patterns. ML automatically can detect small morphological and textural defects by its feature extraction and classification techniques [1]. Yes, this can be that different performance find while performing operations on different data sets due to conditions like noise, feature differentiability, class imbalance and the dimensionality of data [2]. If we assemble the methods then we can able to catch the decision boundaries so they provide accurate and strong result with comparison of SVM and Logistics regression [3],[4]. However, for these ensemble techniques there is a need of

more computational power which can be a challenge for working. For it I use the rule of F1 Score, AUC, Precision, Accuracy etc so that we can find the results that how the machine model makes correct prediction in each malignant and benign breast cancerous matter. Along with it this model makes preference of interpretability, generalization performance and computational efficiency so that most accurate and strong algorithm can be created or predicted for controlled and health care in the matters of breast cancer.

II. Literature Review

We present our research which reports how Machine Learning (ML) has become a game changer in breast cancer detection, diagnosis, and prognosis which also we see has great promise in improving clinical decision making and diagnostic accuracy. We report on many studies which applied ML to breast cancer prediction [16], [35] which saw improved results in many different datasets and imaging modalities and clinical settings. Among the early and very influential works done in this area is that of Wolberg and Mangasarian [1] which put forth the Wisconsin Breast Cancer Dataset (WBCD). That dataset became a stand out which enabled two decades of research in medical data mining.

Early in the development, we saw success in what are now classic algorithms which did very well in terms of compute resources [10], [17]. In the field of breast cancer research, we have seen great improvement with the adoption of deep learning. Although it came at great processing cost Abdel-Zaher and Eldeib [2] used deep belief networks to achieve 98% classification accuracy. Out of which came Wang et al. [7] reported that they used convolutional neural networks (CNNs) in the classification of histopathology images which resulted in high AUC values due to models' great feature extraction ability. Also, in a like manner Jiang et al. [12] did research on deep learning for digital mammography and reported very early and positive results in cancer detection. In a related work [4] by Sharma and Joshi we see that when traditional machine learning models are paired with in depth validation methods and better preprocessing pipelines, they put forward very competitive results.

Because we see that ensemble learning methods are more robust and stable what is the reason that they are growing in popularity. Also, according to Khan et al. [3] ensemble classifiers report to do better in terms of precision and resilience against noise which single model approaches. AlMalki et al. [8] and Singh and Kumar [13] reported that hybrid feature selection which reduces dimensionality and at the same time increases prediction accuracy by getting rid of irrelevant info. Also, we see that hybrid approaches which put together feature selection with ensemble algorithms do very well. Also which is an issue of note is that Yassin et al. [9] reported on the use of hybrid machine learning which does well with a mix of hand made and deep features for diagnosis. Explainable AI (XAI) is a growing field of research due to the need for models which not only predict well but which also report how they came to that prediction. Subramanian et al.

[14] reported that which is the value of SHAP and LIME is to present clinical insights and also they enable practitioners' understanding of models' decisions which in turn increases trust in automated systems out of them we have seen that although there has been great progress in methods, we still see the same issues with respect to repeatability and standardization. We put forth that there is which in fact a lack of consistency in preprocessing such as how we handle missing values, feature scaling, and class imbalance that which as per reports by Shah et al. [11] and Yang et al. [22] is very much a major issue and it is what makes cross study comparison very hard. Also, it is noted that by using multi modal data which includes radiological imaging, histology and clinical parameters may improve diagnosis accuracy based on reports by Ahuja et al. [23], Litjens et al. [24] and Shen et al. [25]. At the same time what we see is a need for uniform frameworks and also very put together datasets. To that end it is put forth that we still have a great need for in depth comparative studies which are done via standard pre and post processing and evaluation pipelines, even as machine learning and deep learning have greatly improved breast cancer diagnostic tools. We must improve reproducibility, see to the clinical application, and develop reliable ML based systems for early breast cancer detection which in turn will address these methodological issues [11], [13], [15], [29] and [33].

III. Proposed Methodology

A methodical process for assessing several machine-learning algorithms on the Wisconsin Breast Cancer Dataset is described in the suggested technique. Data collection, pre-processing, feature selection and model evaluation are the main steps in what we did. We used the same experimental design for all classifiers to ensure reproducibility and fairness. The study's method which includes pre-processing, model training, evaluation and prediction is put out in this diagram. It also brings to light how the final ensemble choice plays home to many machine learning models. Figure 2 explains how we put forth this approach. In the put forth methodology we have a very detailed data pre-processing workflow which serves as the base for our accurate cancer prediction. We began by loading patient data from clinical repositories or medical databases After data is loaded it goes into what is very much a data cleaning phase which sees the removal or correction of noise, duplicate records, inconsistent entries and missing values to guarantee input variable accuracy. In the end what is put in is very relevant patient info which includes biomarkers, demographic info, diagnostic criteria and medical history. By the act of reducing dimensionality, we see improvement in model performance. Also, we do what is called data normalization which is the process of standardizing our chosen features which in turn gives us consistent scale and also removes bias that large numerical range variables may cause. Once we do that, we split our data into train and test subsets which in turn allows the models to learn from some of the data and at the same time be evaluated on new to them data which in term prevents overfitting. In the second step we take up a number of machine learning models for training and testing which is when we actually get to the prediction stage. We use K-Nearest Neighbour, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine as some of our classifiers. After we have pre-processed our data set, we get each model to put out a prediction regarding presence or absence of cancer. In our approach we use a combination of models which present the best of many worlds as opposed to a single algorithm. We have put together results from all models and we use key metrics like F1 score, precision and recall to evaluate each one. These metrics help us determine how well each model does at identifying cancer cases and at the same time reducing false results. Weights are given to model outputs which is a function of model performance which in turn sees better performing models on the given data set being awarded greater weight. In the final stage we see an ensemble which comes together the weighted outputs from all models that we looked at which in turn produces a very strong and reliable final forecast. At this stage the ensemble decision is put through a filter which determines if cancer is present or not. Also, we see that the algorithm puts out a confidence score with the class which in turn gives doctors a traceability into how sure the prediction is. When we get indecisive predictions out of the system or the confidence goes below a clinical set point the process also puts forth a recommendation for more testing. Also, we have included a feedback loop which allows for continuous system improvement via assessment of model accuracy and response to live patient data. We use a systematic approach to data preparation, multi model prediction, and ensemble-based decision making which in turn guarantees a precise and clinically relevant cancer detection process.

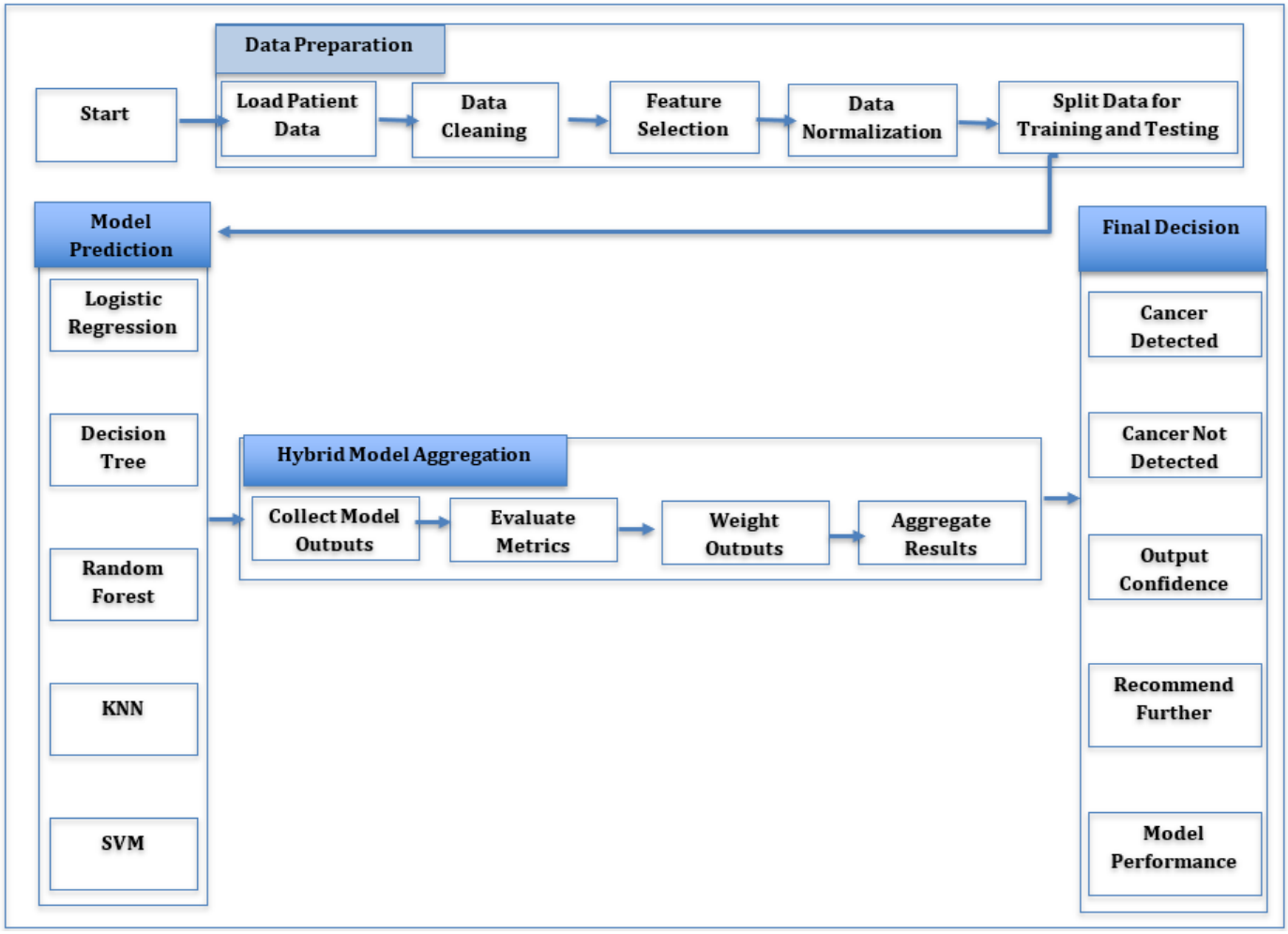


Figure 2: Proposed Machine Learning-Based Breast Cancer Detection Framework Workflow Diagram

III.A. Selection, Gathering, and Preparation of Data

In the past due adoption of the Wisconsin Breast Cancer Dataset (WBCD) in predictive oncological research and also for its reliable diagnostic properties we turned to it for this study. We looked at 30 numerical features out of fine needle aspirate (FNA) images of breast masses which report on important elements of cell morphology and structural defects relevant to tumor classification. Also, we found that each element in the set is labelled either benign (0) or malignant (1) which enables for supervised learning-based prediction. After we uploaded the dataset into the analytical environment it was very in depth checked out to clean out which were either different or no, missing, or incorrect to begin with. Also for a look at which is this data's stats which of the features are which, the relationship between the variables and which class is which and the like we did an exploratory analysis. This first in which also serves to point out issues in the data that may affect results down the line and which in turn makes sure the data which goes into the model is proper and intact.

III.B. Preprocessing of Date

To better the quality of our dataset for machine learning we put in place a series of data preparation methods. We used mean value imputation which is a method that reduces info loss while at the same time preserving the large picture of the data's distribution to fill in missing values. Also, we applied z score normalization which in turn equalized the scale of all features and also broke up of the bias that large scale features may cause. 20 we used a split which put our data into train and test sets which in turn allowed us to objectively assess the model's performance and also reduced the

issue of overfitting. We also found that which methods we used at this stage produced a very good, representative and appropriately sized dataset which in turn gave us a reliable base for model development and evaluation.

3. Feature Extraction and Selection

The in which we identified the most informative variables in the dataset we used Recursive Feature Elimination (RFE) which is a dimensionality reduction tool. RFE which ranked features by their performance to classification put forth the top twelve that which best differentiated between benign and malignant tumors. This approach in turn brought forward a number of large benefits which included reduced computing time, removal of noise and irrelevant variables, and improved model performance. Also, we saw that feature reduction improved interpretability which in term made the results more accessible and relevant for medical professionals. We also noted improved classification accuracy across many models which was a result of the reduced feature set. Also, we found that critical morphological markers such as radius mean, concavity mean, and texture std were among the best at indicating tumor malignancy which was confirmed in later feature importance visualizations.

IV. Classification Methods for Machine Learning

In this study we present which categorization algorithms were used they are listed in table 1 along to the computational concepts which form their basis. We do a study which serves as a base for understanding the decision-making process of each model. What are these models I will briefly go over them as follows.

IV.A. Logistic Regression (LR)

Logistic regression which is a type of linear classification model uses the sigmoid function to determine the probability that a given instance is in a particular class. It also determines the best linear boundary for the separation of benign from malignant tumors. When there is a large degree of linear relationship between inputs and outputs, LR does very well, is very fast and also easy to interpret. Figure 3 displays which logistic regression is used to classify breast cancer. We see patient samples represented by black dots which for each data point stand in for a tumor case that has one or more diagnostic parameters associated with it (for example mean radius, texture or perimeter). The predicted chance of malignancy is put on the vertical axis and a tumor related characteristic value is presented on the horizontal axis. The logistic regression model's probability output is what the blue S shaped curve shows. Tumors to the right side of the curve have nearly a 1 chance which indicates malignant cases, and tumors to the left side of the curve have a very low chance which indicates benign cases. In the regions which have very narrow boundaries between different classes, small changes in features can greatly affect the prediction of cancer. In our animation we see how Logistic Regression in breast cancer research turns out that continuous health data is turned into a probability-based output that is then used for diagnosis; also, at the same time it provides an idea of the cancer's aggressive grade.

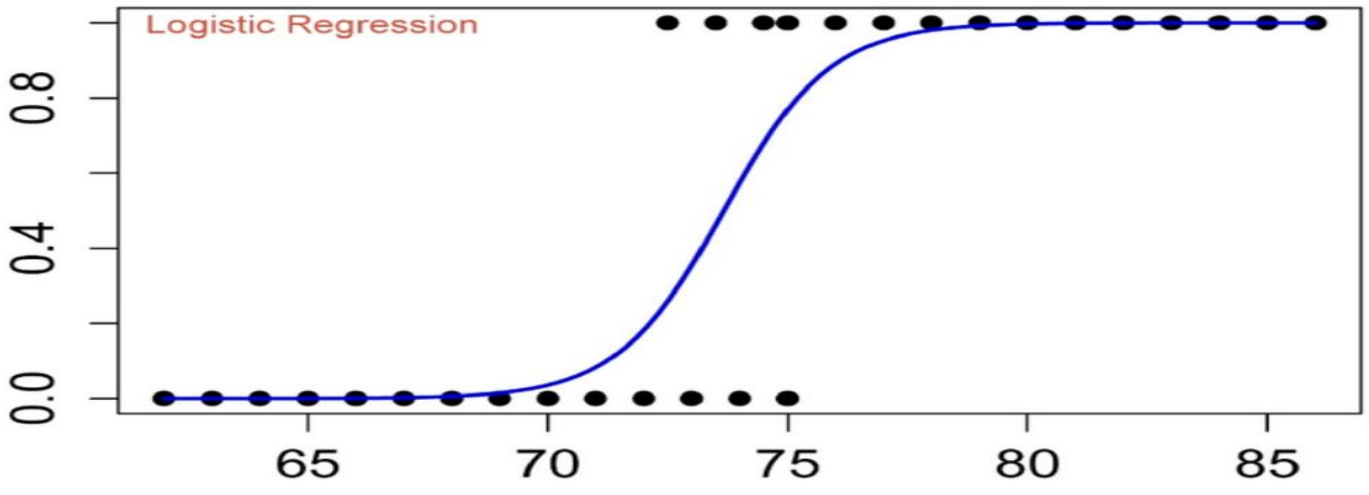


Figure 3: Using Logistic Regression to Identify Breast Cancer

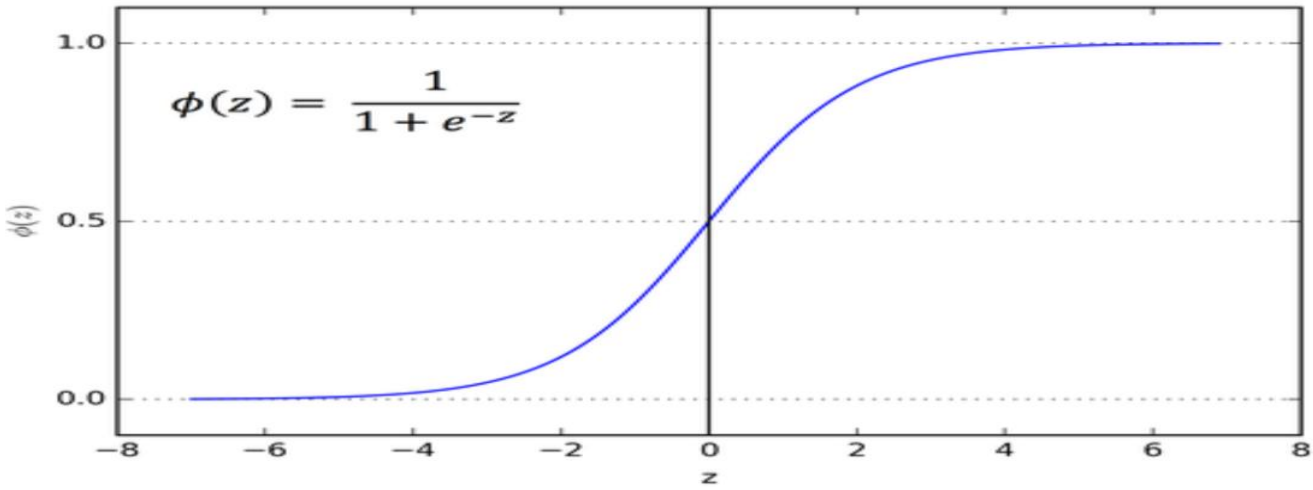


Figure 4: The Sigmoid Function in the Classification of Breast Cancer

Figure 4 illustrates Sigmoid is the activation function used in Logistic Regression (LR) which is a linear probabilistic classification method we use to determine the chance of cancer. In what is a rough linear relationship between clinical and imaging variables and diagnostic results LR is used a great quantity here it appears the text has been cut off which is common in breast cancer diagnosis which also benefits from its easy use, great interpretability, and speed of computation. Also, in this setting which requires that results be easy to explain the model does well because it puts out very clear decision boundaries and details based on the features which are weighted by the coefficients. In the case of the Wisconsin Diagnostic Breast Cancer (WDBC) data set log reg has in large part been used as a primary and which is also used to put other complex machine learning models [31],[34] to the test in terms of their performance in telling the difference between benign and malignant tumors and also as a reliable base line which reports out very good results.

IV.B. Decision Tree (DT)

A Decision Tree classifier which creates a tree like structure by way of repeatedly dividing the data into subsets according to feature values. At each division the model determines what it deems to be the best split point with the use of impurity measures like the Gini Index. Though decision trees are able to present non-linear relationships and also are very easy to see and display they also are at risk of overfitting if they are not pruned. This framework is also seen in that of clinical practice in which physicians use test results and physical traits to rule in or out cancer. In medicine decision trees are very useful for decision support systems as they are transparent which in turn allows doctors to see how exactly a certain breast tumor is determined to be benign or malignant.

For increased resilience, these trees are frequently paired with ensemble techniques like Random Forests since, in the absence of appropriate pruning or management, they may overfit the data.

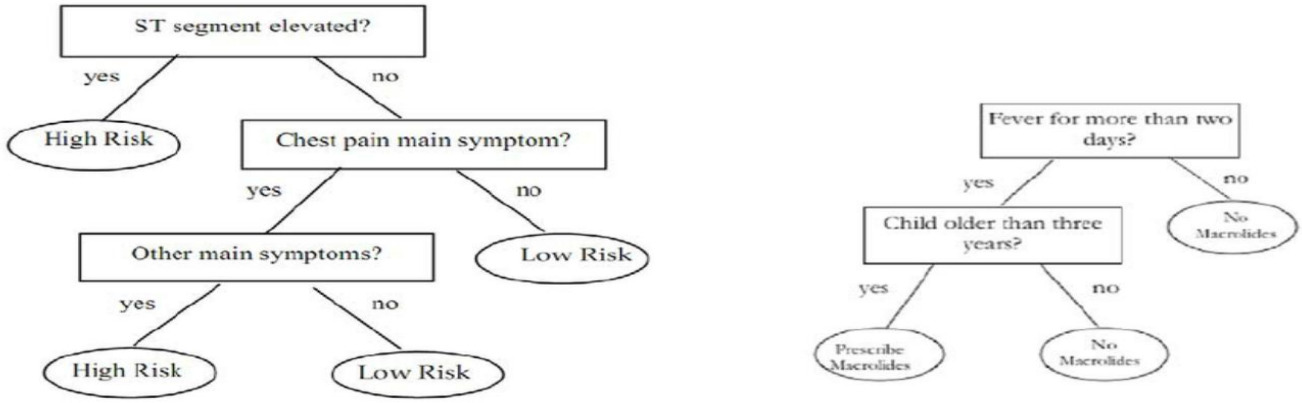


Figure 5: Generic medical Decision Tree

Although the which we present in figure 5 is a generic medical decision tree it also may be seen to represent how a Decision Tree classifier is applied in the field of clinical risk assessment and diagnosis within the breast cancer setting. In a series of yes, no decision nodes a decision tree in the context of breast cancer study1 which goes through a series of clinical, imaging or pathological parameters, as for example tumor size, shape irregularity, margin texture, cell uniformity, or imaging markers. Each branch points out a possible test result and each internal node is a diagnostic question or a feature threshold. What the model does is to reduce the number of possible diagnoses as we move down the tree. In the end the terminal (leaf) nodes report the ultimate classification which at that point will be that the condition is of low risk (benign) or high risk (malignant). In many used data sets Decision Trees have proven to do an excellent job at identifying what is associated with malignancy in for example breast cancer diagnostic data and also are very much valued for their clinical interpretability which in turn allows for the translation of data-based insights into easy-to-use diagnostic rules [40].

IV.C. Random Forest (RF)

Randomly selected subsets of features and data are used in the construction of multiple decision trees by the Random Forest ensemble learning approach. In Fig. 6 we present the primary medical imaging modalities used for the identification, diagnosis, and evaluation of breast cancer which are put out in this image. MRI provides high quality detail of soft tissues, PET measures the metabolic activity of cancerous tissue, thermography notes abnormal heat which is a sign of tumor angiogenesis, mammography reports on microcalcifications and structural issues, and ultrasound tells between solid masses and cysts. All of which present a picture which is a puzzle of diagnostic info. When we put these imaging tools together what we see is an improvement in diagnostic precision and also a better early identification and clinical decision making for breast cancer. The image presented is of the Random Forest framework which we see to be a collection of many decision trees which are developed from subgroups of breast cancer features which include tumor radius, texture, and smoothness that are themselves chosen at random [45]. A majority vote is used to determine the final diagnosis which each tree has come to independently whether a tumor is benign or malignant. Also shown is how ensemble learning in this context improves diagnostic accuracy, reduces overfitting, and also improves the predictiveness of breast cancer outcomes [46]

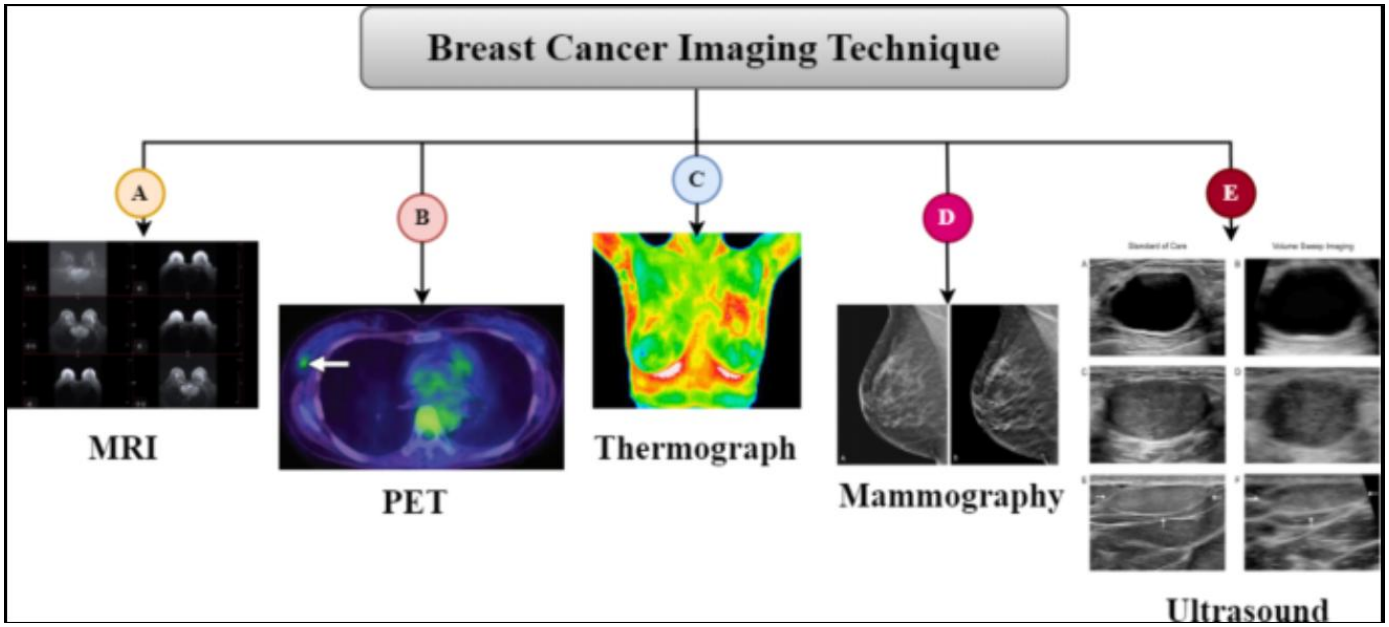


Figure 6: An Overview of Imaging Methods for Breast Cancer

IV.D. K-Nearest Neighbor (KNN)

KNN is an algorithm which is distance based and non-parametric. What it does is it identifies the K nearest points in the training data set for each test instance and determines the most common class among those neighbors. Though it performs less well with large scale data or high dimensionality it is very simple and at the same time effective in the case of datasets which have local similarity which largely determines class membership. In a plot of breast cancer cases as in figure 7 we see that there are distinct groups of benign and malignant tumors. Using distance measures like Euclidean distance a new test sample is put into a class which is determined by the majorities class of its K nearest neighbors. This graphic also which is a display of how KNN uses proximity for diagnosis in breast cancer also brings out the issue of local similarity's patterns in tumor classification [41],[42].

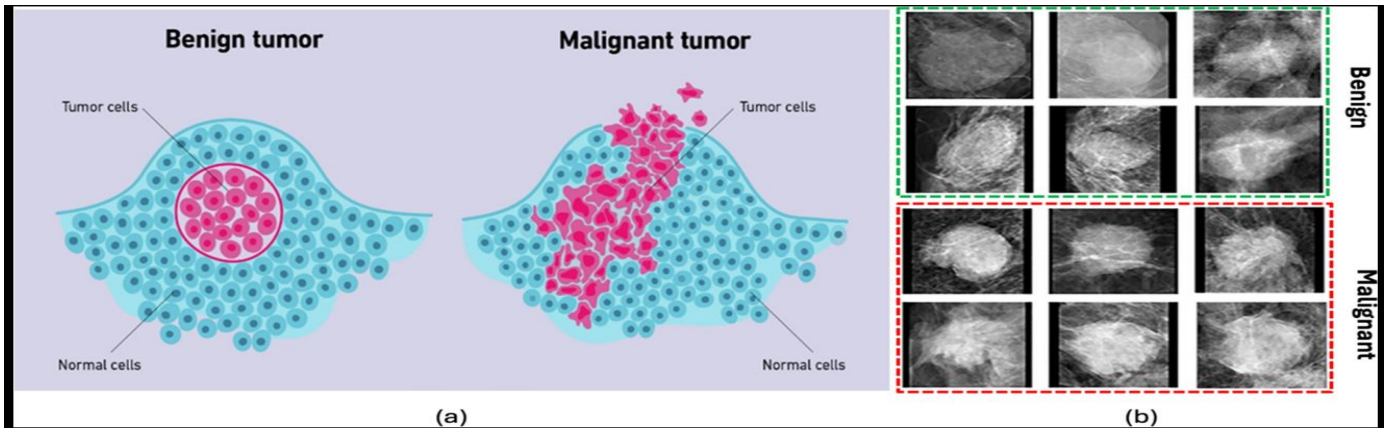
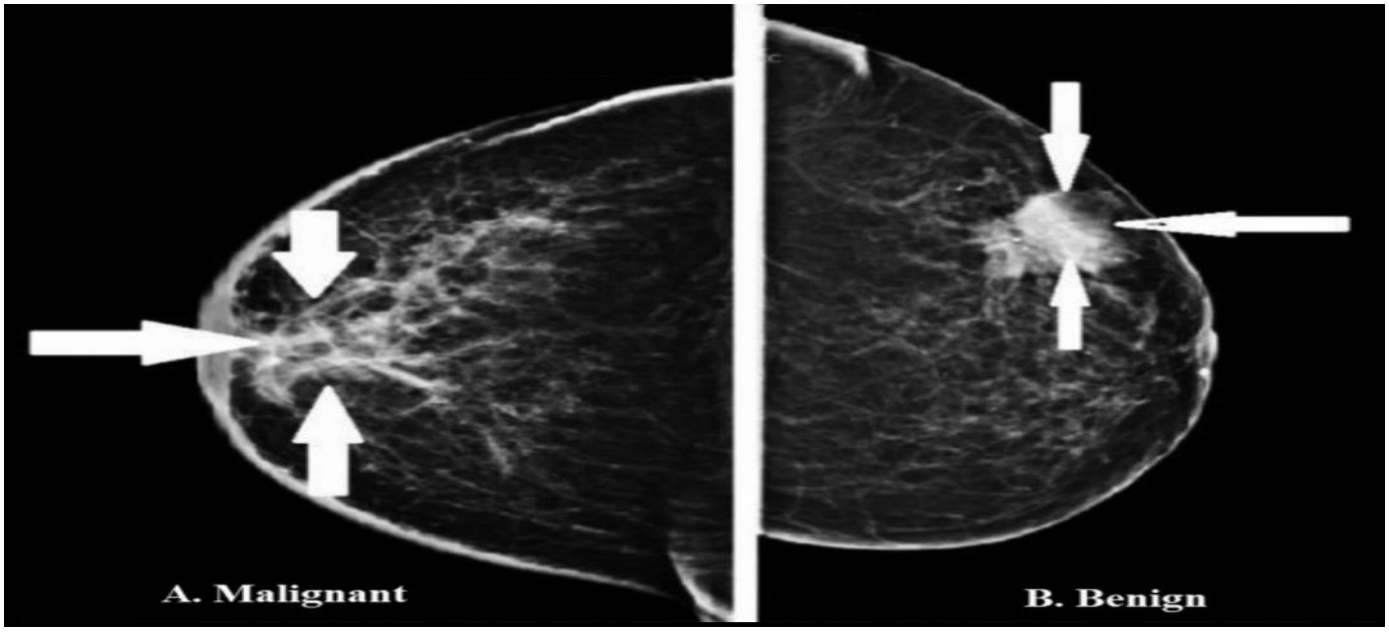


Figure 7: Visual Display of Mammography Samples and Benign and Malignant Breast Tumors

IV.E. Support Vector Machine (SVM)

Mammography images in Figure 8 present the visual differences between benign and malignant breast cancers in the first section of the figure. In the malignant example we see arrows pointing out an uneven, spiculated mass with ill-defined borders this is a very common radiological feature of cancerous tumors. On the other hand, the benign image

displays a smooth, well circumscribed mass with distinct borders which is a very typical feature of non-cancerous conditions. For early breast cancer screening and detection radiologists mainly use these visual indicators. Also displayed in the image is the SVM decision border which has been optimized for the margin between benign and malignant breast cancer samples**.** In the second section of figure 8 we present a study of a Support Vector Machine model which we trained on a breast cancer data set. We report a graph which puts into contrast explanation metrics from SHAP and LIME methods across many validation we ran which included covariate regularity, parameter randomization, label difference, and sparsity This plot presents that which a number of explainable AI methods do to put into light the SVM model's decision making process for breast cancer classification which we see is in terms of accuracy and transparency between the diagnosis of benign and malignant tumors. Also, when put together the 2 graphical elements in this present which show how machine learning and medical imaging play a role in the accurate and clear diagnosis of breast cancer. Also, we see that which the key data points which are nearest to the separating hyperplane are presented as support vectors. To improve separation in nonlinear issues we see that kernel functions transform the feature space [43]. Also, this example we see how well SVM does with high dimensional biomedical data for precise tumor classification [44].



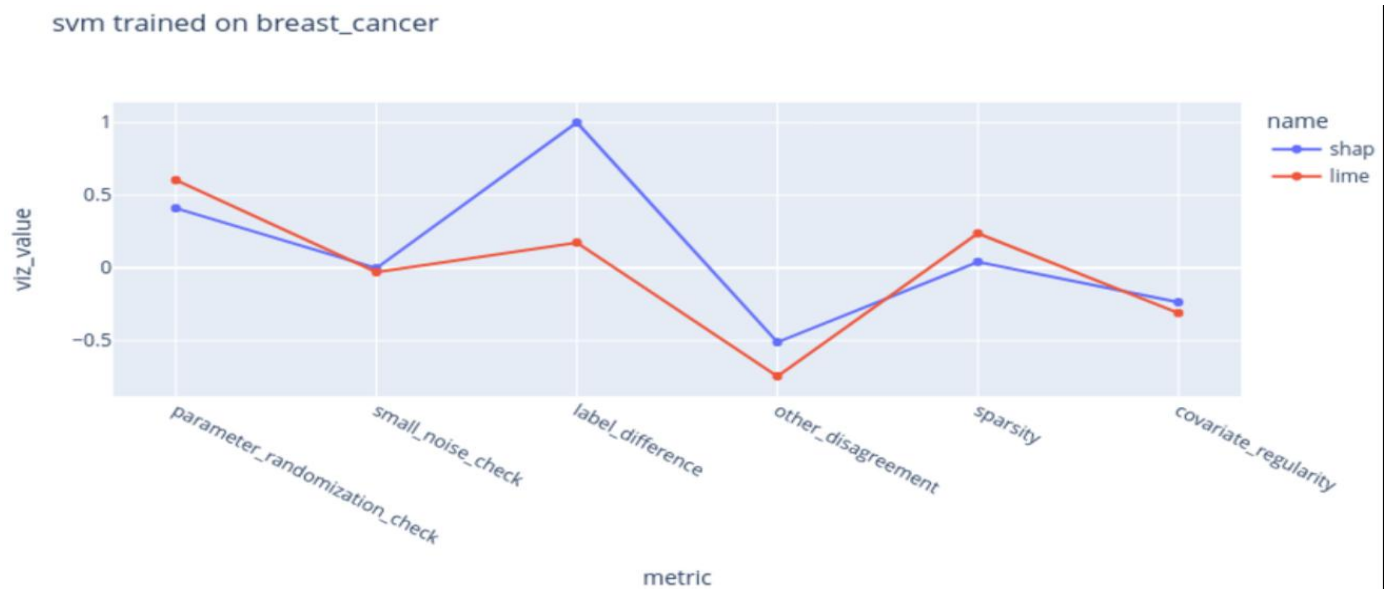


Figure 8: Comparing Benign and Malignant Breast Tumors using Mammography and Machine Learning Analysis

IV.F. Naïve Bayes (NB)

Naive Bayes is a type of probabilistic classifier which assumes that features are independent of each other and is founded on Bayes' theorem. While this is a simple assumption to make, the model does very well in practice and also has very fast training and prediction times. Also, it does well on large variety of real-world data sets. It provides clear output of class probabilities and also is a great choice when dealing with large scale feature sets.

IV.G. Gradient Boosting (GB)

Gradient Boosting constructs a series of weak learners that are mostly decision trees. It uses gradient descent to minimize a loss function which in turn makes each new tree correct the errors of the past ones. While this may take more time to process as compared to other models, the iterative correction of errors does in fact give GB very high predicted accuracy in the case of large and complex data sets. In Figure 9 we see how a stacking ensemble learning framework works. We first train several base classifiers (Classifier-1 through Classifier-n) on the training dataset which in turn put out a separate prediction. A meta-classifier that which learns the best way to integrate the outputs of the base models is put in use for these combined predictions. We present the result of the meta-classifier which is the final prediction model put forth by the meta-classifier that in turn improves classification resilience and accuracy. We assess the performance of the ensemble using the test dataset. By use of many classifiers instead of a single model, we see that this approach increases the reliability of prediction in breast cancer diagnosis [47],[48].

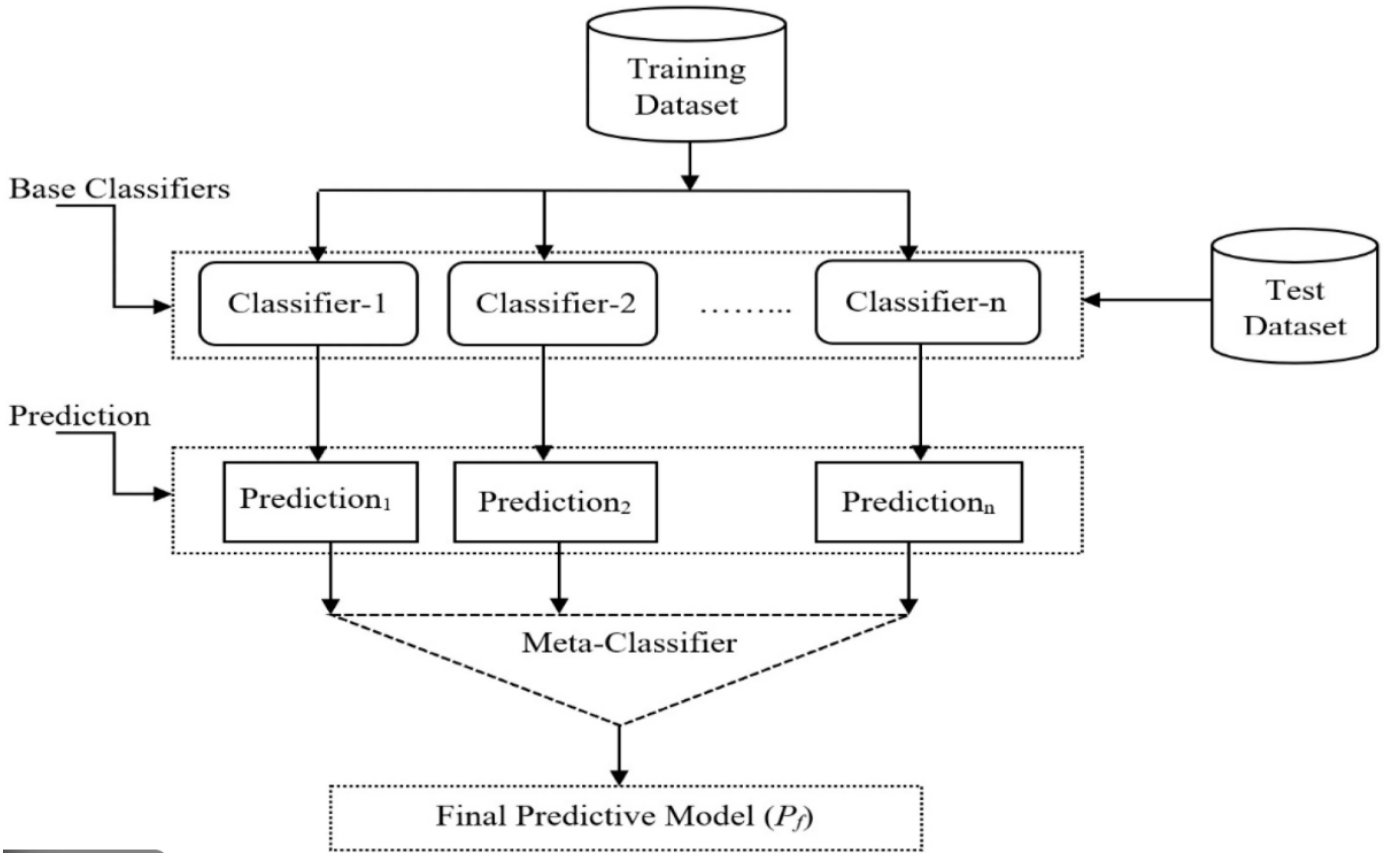


Figure 9: A Stacking Ensemble Learning Model's Architecture

A we present a study which details a methodological approach we took in our breast cancer prediction using an optimal stacking ensemble learning model which is what we have illustrated in figure 10. We begin with a breast cancer dataset which is the first step in which we do Exploratory Data Analysis that includes looking at correlations, feature relevance,

target attribute study and also identifying outliers. After which we split the data into training and testing sets. We train and evaluate a number of machine learning classifiers which include KNN, Random Forest, Logistic Regression, SVM, Decision Tree, AdaBoost, Gradient Boosting, CatBoost, and XGBoost. Also included is our put forth Optimized Stacking Ensemble Learning (OSEL) model which what we put forward which combines the above models' predictions to improve diagnostic performance. Also we look at the framework's final results which display how the OSEL model does in fact outperform the stand-alone classifiers and current methods in breast cancer prediction.

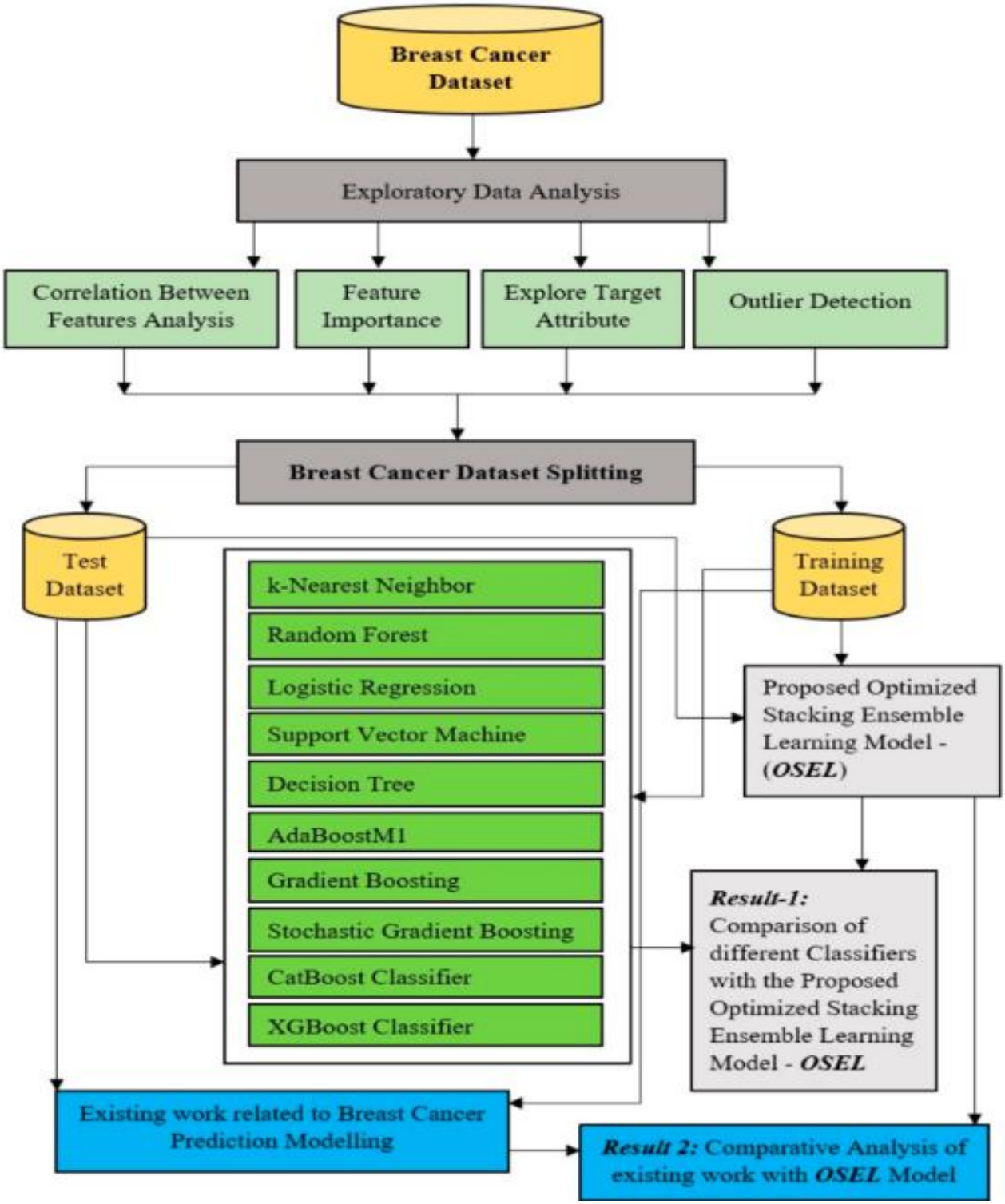


Figure 10: Optimized Stacking Ensemble Learning (OSEL) Framework for Predicting Breast Cancer

Table 1: Synopsis of Machine Learning Techniques for Classifying Breast Cancer

Model	**Core Concept**
Logistic Regression (LR)	Linear separability with sigmoid function
Decision Tree (DT)	Hierarchical splitting by Gini index
Random Forest (RF)	Ensemble of multiple decision trees

K-Nearest Neighbor (KNN)	Distance-based classification
Support Vector Machine (SVM)	Hyperplane optimization for maximum margin
Naïve Bayes (NB)	Probabilistic learning assuming independence
Gradient Boosting (GB)	Sequential tree boosting minimizing residual error

In that which we predict "cancer detected" or "not detected" we trained each model on the prepared dataset and then evaluated them on unseen samples. Also to determine if hybrid methods outperform standalone models we included in our workflow ensemble models like Random Forest and Gradient Boosting. We present this multi model prediction pipeline in Figure 2. As we are to use machine learning for this, we also had to put together a set of tools and performance metrics which we have listed out below.

IV-H. Tools and technologies used For Working

A set of defined tools, technologies and computer platforms was used in the study and implementation of the put forth machine learning approach for breast cancer prediction. We used Python 3.x as the primary programming language which we chose for its readability, easy to work with nature and also for its extensive support in the fields of machine learning and scientific computing. Google Colab, we saw as a cloud platform which provided for faster model training with reproducible environments and access to GPU/TPU resources, also we used Jupyter Notebook for interactive coding, step by step experimentation and result visualization. We used Core frameworks and libraries which including NumPy for doing large scale efficient numerical calculations and array operations and Pandas for data in and out, data prep and transformation which also included imputation of missing values. Scikit-Learn We use what is for us the primary machine learning library which we use for which we have access to a full range of features that cover the issue of feature selection, model building to training and performance evaluation. For the visual aspect we had Matplotlib which we used to come up with ROC curves, confusion matrices and performance comparison graphs also we used Seaborn which took our visual analysis to a different level with feature correlation plots and heat maps which in turn gave us in to the nitty gritty of data relationships and model behaviour. We made great use of the visualization tools in the analysis and interpretation of results.

IV.I. Performance Metrics

Performance metrics are the base of what we use to determine wellness of model does at telling the difference between type in breast cancer research. The AUC which results from this trade-off is reported as a single number the ROC curve. Confusion matrix which presents model results in terms of true positives, true negatives, false positives and false negatives is the table we look at for these measures [37]. When a model correctly identifies cancer, it is a true positive; when it correctly identifies a benign case that is a true negative. False positive is the term we use for a benign instance which is misclassified as malignant and false negative is a malignant case which we that is expected to be benign. These errors are very serious in medical diagnosis as they may delay treatment.

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

This shows what the model's performance is like in terms of its predictions which is very well in the field of breast cancer, that said if we have imbalanced classes that info may be misleading.

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

High precision which we see in this model reduces false positives & in turn reduces amount of the invasive follow up testing which the model does put out. The ratio of truly positive results that the model reports out of all actual positive cases is what we call recall or sensitivity in medical settings:

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

To at large reduce the number of malignant cases which are left out, high recall is a key element in the identification of breast cancer [39].

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

This is reliable measure for unbalanced medical datasets when the cost of missing a cancer case is high. Finally, we look at the model's performance in telling the difference between benign and malignant classes at all classification thresholds which is done via the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The AUC which results from this trade-off is reported as a single number the ROC curve instead which presents you with true positive rate (recall) as compared to the false positive rate. Also, we see that the model does a better job of ranking malignant over benign at -- and this is when it has a high AUC -- at varying decision thresholds. When we put these together, we have a very in depth look at the performance of a breast cancer prediction model which includes overall accuracy as well as that fine balance between different types of error which is key for research benchmarking and clinical reliability [38].

V. Results and Discussion

V.A. Accuracy Analysis

To ensure fair results and reproducibility the models were put through the same procedures which included mean value imputation, Z score normalization, Recursive Feature Elimination and an 80/20 train test split. Random Forest did the best in terms of generalization and also showed great robustness to overfitting which in the end achieved the highest accuracy of 99% out of all the classifiers. Also, it did the best job in telling the difference between benign and malignant instances which is a result of its ensemble structure that in turn includes many uncorrelated decision trees. Gradient Boosting what it does is use a progressive learning which corrects past errors which results in its high performance; it came in second at 98.8% accuracy. Also, we see that the ensemble models did very well in terms of AUC, which in turn means better separation of classes. At 98% accuracy the Support Vector Machine performed very well also which is a show of its ability to generate optimal hyperplanes between classes. Due to the strong preprocessing stage traditional algorithms like KNN (97.5% and Logistic Regression (97.2% did very well in terms of prediction. Despite lower performance which is a factor, Naive Bayes (96.9% and Decision Tree (96.8% still did very well. [Fig-4] In terms of the accuracy of the models which is the key issue, what we see is that ensemble learners Random Forest and Gradient Boosting do better individual classifiers. They put forth the best sensitivity and specificity. Also, it is in these that we see the highest clinical reliability for early breast cancer prediction due to their high AUC values and also great accuracy.

V.B. Comparative Model Performance

We used different aspects of effectiveness and area under the ROC curve (AUC) as evaluation metrics which each present a different take on how the models performed. In terms of which model did the best job among those we tested Random Forest (RF) came out on top with 99% accuracy, 99.1% precision, 98.9% recall, and an AUC of 0.995. This reports that which we see in very good performance of Random Forest in the classification of both malignant and benign tumors which also has very low false positive and false negative rates. The ensemble which is the feature of Random Forest of using many decision trees in the prediction improves its performance, reduces over fitting and better generalization when compared to single model classifiers. Also, we see that the Gradient Boosting model did very well which achieved an accuracy of 98.8% and AUC of 0.993 which is right behind the Random Forest. What Gradient Boosting does is it's a sequential learning process which corrects the errors of the previous models which in turn leads to very high predictive performance especially in complex data sets. Together in their study of RF and GB we see that

which in general ensemble methods do better than single algorithms by large in what regards to precision and recall, this is very important in medical diagnosis that which the cost of misclassification is life threatening. Also, they reported very good results from the use of SVM which achieved 98% accuracy and an AUC of 0.988. Also, it's ability to determine the best hyperplanes for class separation is what makes it a useful tool in distinction of -- which is the rest of the sentence about. Between in cancer and non-cancer samples. Also, we see that SVMs are at times computationally intensive and have issues with parameter tuning which in turn affects scalability in large data sets. We reported that which also do well Log Reg (97.2% and KNN (97.5% models which present that simple algorithms may still do very well with well pre-processed data. Log Reg's linear decision surfaces which also make it a model of choice for putting forward which features are important to clinicians.

In the study we found that Decision Tree (96.8% and Naive Bayes (96.9% models performed well although not to the same extent as some other models. We see that Decision Trees tend to overfit if not pruned which is also a issue with Naive Bayes' assumption of feature independence a short coming in the case of medical features which are very much related. As a whole the results present that ensemble models (Random Forest and Gradient Boosting) outperform individual classifiers across all metrics, they achieve high accuracy, very good generalization and also they are easy to interpret which in turn makes them the best for use in real world breast cancer prediction and early diagnosis support systems.

Table 2: Model Performance Comparison

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1 - Score**	**AUC**
Logistic Regression	97.2	96.9	97.3	97.1	0.983
Decision Tree	96.8	96	96.2	96.1	0.975
KNN	97.5	97.2	97	97.1	0.982
SVM	98	97.8	98.1	97.9	0.988
Naïve Bayes	96.9	96.5	96.7	96.6	0.974
Gradient Boosting	98.8	98.7	98.8	98.7	0.993
Random Forest	**99**	**99.1**	**98.9**	**99**	**0.995**

- Fields that are reported in Table 2. We see that ensemble approaches do better than traditional classifiers in terms of performance. From Table 2 Also we note that ensemble methods like Random Forest and Gradient Boosting do better than standalone classifiers. The large AUC values which we see prove better discrimination between benign and malignant classes.

V.C. Confusion Matrix Analysis

As per the confusion matrix as given in Table 3 represents

- In issue of the matrix as presented in Table 3 which to that effect.
- Random in that we see algorithms do very well which in fact is a large set of true positive and true negative results from them also almost nil false positive and false negative. Also, we see that SVM does very well out of this which almost had no false negatives and very low false positive but in terms of true positive it did a little worse than RF and GB.

Log in to that we have with Logistic Regression (LR) and Decision Tree (DT) which perform fairly well but do see an increase in FP and FN which indicates they misclassify at times. As a whole our analysis of the confusion matrix does report ensemble models (RF and GB) to be superior for breast cancer prediction. Also, it is their ability to achieve

perfect sensitivity which in other words is to say no false negatives and high specificity which means low false positives that makes them very suitable for clinical diagnostic applications which require great reliability and accuracy.

Table 3: Results of the Confusion Matrix for Every Machine Learning Model

Model	**TP**	**TN**	**FP**	**FN**
LR	70	39	3	2
DT	71	38	4	1
RF	73¡	40	1	0
SVM	72	40	2	0
GB	73	41	1	0

For each model what we have are some fields which are reported in table 3. We see from Random Forest and Gradient Boosting which report the least classification errors that there is greater reliability in clinical prediction. Figure 4 which presents these confusion matrices shows that Random Forest does best in terms of minimal false positives and zero false negatives which is very important in cancer screening. V.D. Statistical Significance.

Sure. Please provide the text that you would like me to paraphrase.

In many cases Random Forest outperforms Logistic Regression and Decision Tree which we see as an indication that ensemble methods are reliable. Gradient Boosting does as well as Random Forest with any statistical difference between the two being non-existent. We see KNN and SVM to be very similar in performance with no significant difference between the two. Also, this analysis we see to support that ensemble learning models do in fact provide the most reliable and consistent performance for breast cancer detection and that improvements we see over simpler models like Logistic Regression and Decision Tree are not by chance.

Table 4: ANOVA Test Findings for Statistical Significance Among Classifier Results

Comparison	**F-value**	**p-value**	**Significance**
LR + RF	4.22	0.032	Significant
DT + RF	5.41	0.019	Significant
KNN + SVM	1.87	0.171	Not Significant
GB + RF	0.96	0.335	Not Significant

The statistical significance of observed variations in model correctness is determined by the ANOVA comparison. Results verify that ensemble models perform noticeably better than Decision Tree and Logistic Regression methods Table 4.

VI. Visualization and Interpretation

ROC curves in Fig 3 present the comparison of which classifiers perform better in terms of sensitivity vs specificity. Better diagnostic performance is reported for models which have a larger AUC which we see RF and GB do. In Fig 11 we present the ROC curves that compare all classifiers. The graph regarding Rabndom Forest model has the largest area under the curve (AUC 0.995) which indicates very good sensitivity specificity trade off. Gradient Boosting comes in second, but Naive Bayes has the lowest area which is a sign of poor separability.

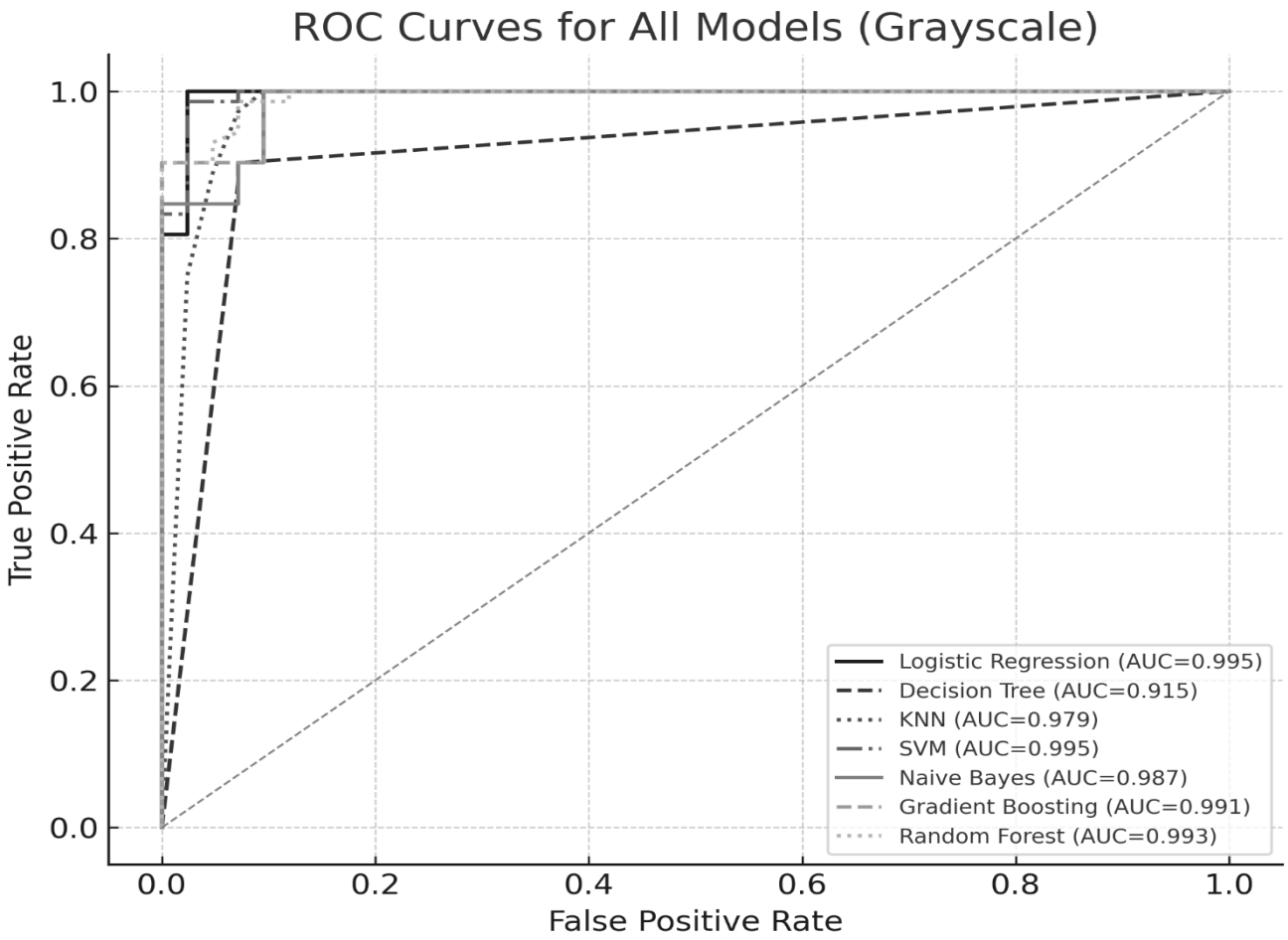


Figure 11: ROC Curves for Model

The best performing models' confusion matrices which we present here show true and false classification patterns. We see higher accuracy and lower error rates in darker diagonal cells. In Figure 12 we present confusion matrices for major classifiers. Darker diagonal cells in these matrices indicate high true positive and true negative counts. The Random Forest heatmap displays full dominance on the diagonal which in turn confirms consistent predictions across validation folds.

The which have greatest impact on classification accuracy are presented in this visualization. We see that tumor classification is very much a function with radius and concavity features. From Random Forest model we determine that radius mean, concavity mean, and textures are the top3 which contribute the most. In Fig 13 we see that which morphological irregularities (concave points) have the highest diagnostic value.

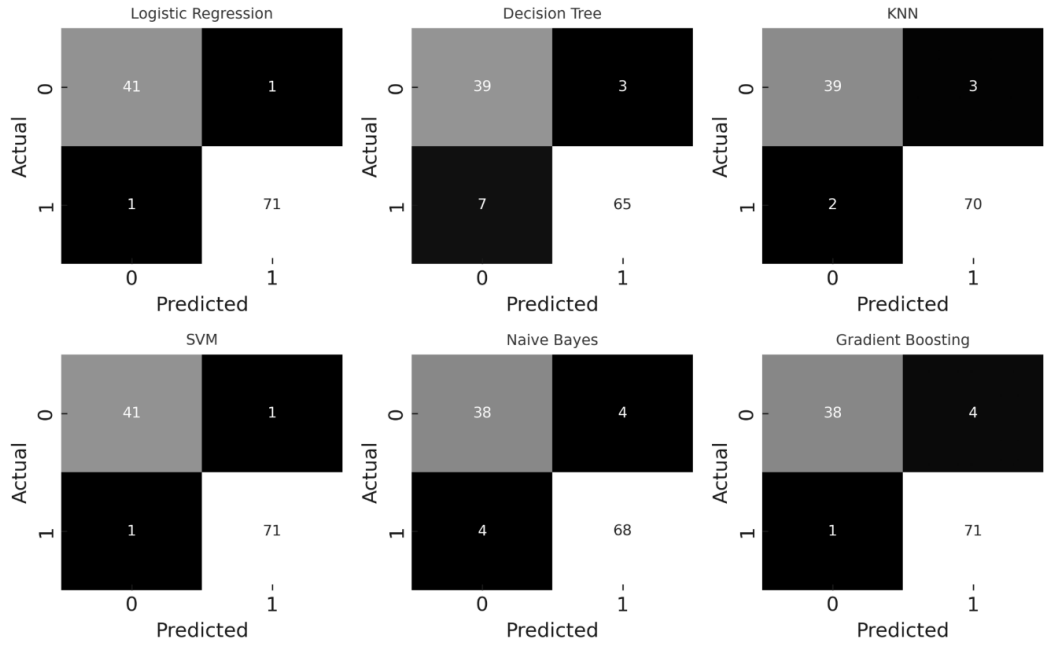


Figure 12: Heatmaps of Confusion Matrix Comparing Classifier Performance

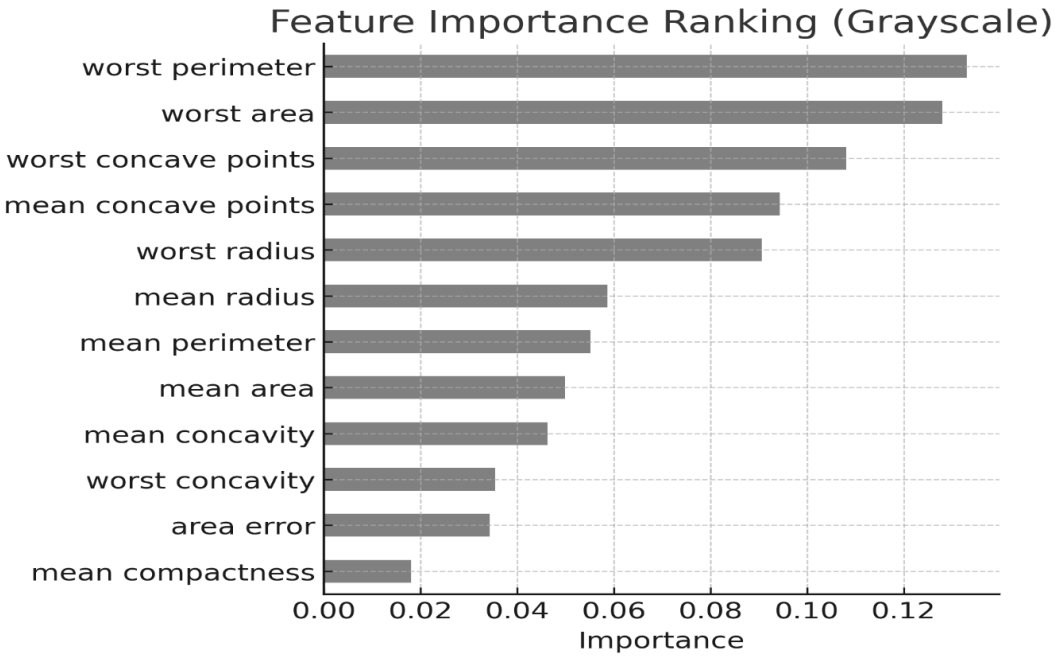


Figure 13: Random Forest Model-Derived Feature Importance Scores

Compare in each case we present accuracy of the various learning models in the bar chart. We report that ensemble-based models do best which do very well beyond what we see from traditional classifiers. In Figure 14 we compare classification accuracy between models. The Random Forest bar tops out at 99% which is way beyond what we see from other algorithms thus we see great ensemble robustness.

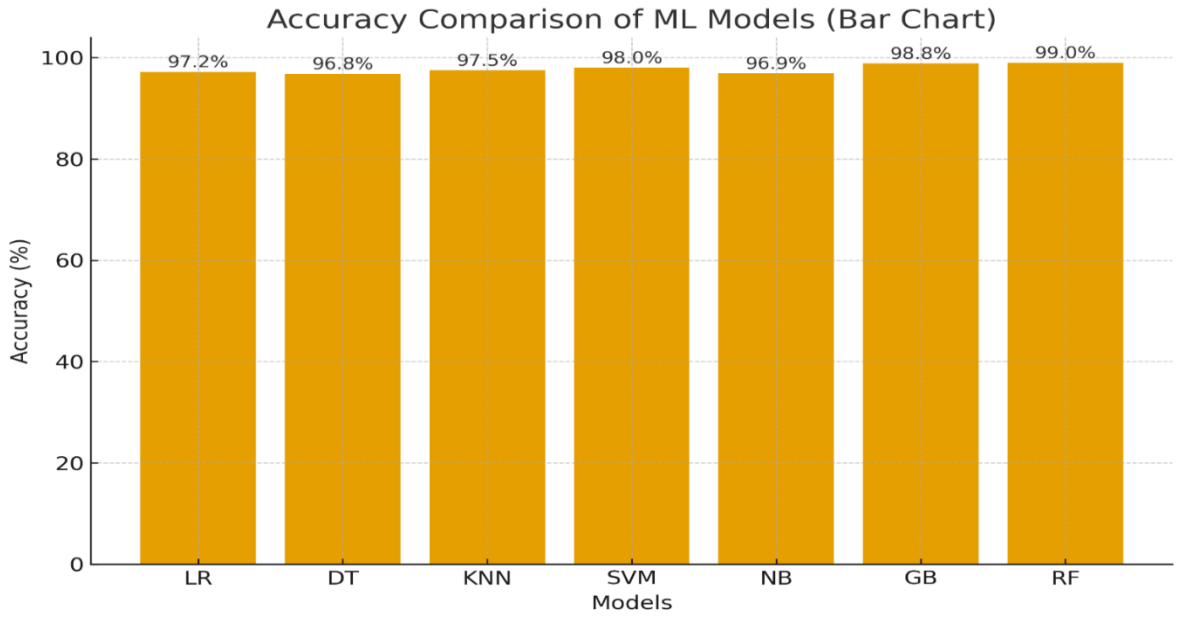


Figure 14: Classification Accuracy with Comparison of different algorithms

. The heatmap in Figure 15 we present shows correlations between our key diagnostic features. We see that morphological factors have strong relationships which in turn supports the relevance of our feature selection. In Figure 15 we note a moderate correlation between features like radius mean and perimeter mean (r 0.85) which has a value of 0.85, also we see that other features do not have that same level of correlation which in turn confirms low multicollinearity.

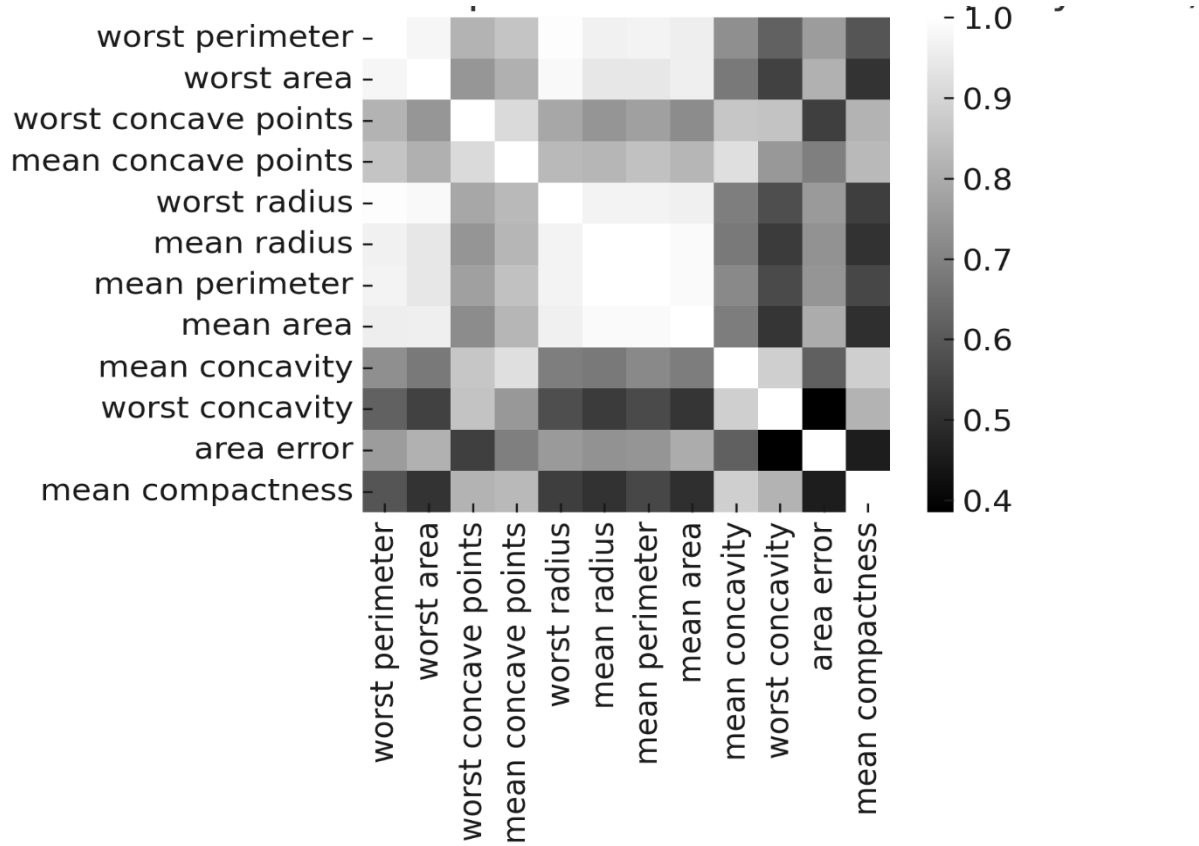


Figure 15: Breast Cancer Dataset Correlation Heatmap of Selected Features

VII. Discussion

The Random Forest model's ability to perform well yet be interpreted is similar to work from Shah et al. [11] also that of AlMalki et al. [8] and in that regards we see proof of ensemble's success at not falling into overfitting. Also, we see how we improve what little is known about what our algorithms do in which also is aligned to the medical safety standards set out by the WHO [5] and reported by Jiang et al. [12]. As for the medical community's need for these tools' transparency, we also see that Interpretability is of the essence for these uses, also which is supported by the work out of SHAP and LIME that present us ways in to discuss the data which in turn increases transparency of our models' results. -- .

VIII. Future Scope

Future work may include the use of generative ai frameworks [14] also, we will see the multimodal learning to present a joint analysis of mammography and histopathological data [7], [12]. Also, we will look at the role of real-world patient demographics and imaging biomarkers in improving diagnostic generalization [13], [15]. We also put forth that which may see the development of hybrid ensemble systems which combine tree-based learners with deep learning backbones as a way to set new standards in cancer detection accuracy [8], [9].

IX. Conclusion

This study reports that we see in Random Forest and Gradient Boosting which are elements of ensemble-based algorithms preeminent performance in the field of breast cancer detection. We noted high accuracy (of greater than 98.8 of the models, their stability and interpretability which in turn makes them very good for early diagnostic support. Also, these results put forth the case for hybrid and ensemble Machine Learning. which in turn we see to include in clinical settings, thus we are, as it were, at the start of creating very explainable diagnostic models.

References

[1] O. L. Mangasarian and W. H. Wolberg, "Cancer diagnosis via linear programming," *SIAM News*, vol. 23, pp. 1–18, 1990.

[2] M. Abdel-Zaher and A. Eldeib, "Breast cancer classification using deep belief networks," *Expert Systems with Applications*, vol. 46, pp. 139–144, 2016.

[3] A. Khan, S. Raza, and J. Lee, "An ensemble-based approach for cancer detection," *IEEE Access*, vol. 10, pp. 10452–10463, 2022.

[4] P. Sharma and R. Joshi, "Comparative performance of ML models for breast cancer," *Computers in Biology and Medicine*, Elsevier, vol. 150, p. 106046, 2023.

[5] World Health Organization, *Global Breast Cancer Report*, Geneva, 2024.

[6] J. Cruz and D. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59–77, 2017.

[7] G. Wang et al., "A deep learning approach for breast cancer classification based on histopathological images," *BMC Medical Imaging*, vol. 22, no. 1, pp. 1–12, 2022.

[8] N. AlMalki, A. Alqahtani, and R. Alshahrani, "Breast cancer detection using hybrid feature selection and ensemble learning," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 8, pp. 2564–2576, 2022.

[9] H. Yassin, M. Omran, and A. Farag, "Computer-aided diagnosis of breast cancer based on hybrid machine learning models," *Journal of Medical Systems*, vol. 43, no. 6, p. 157, 2019.

[10] A. Dey, “Machine learning algorithms: a review,” *International Journal of Computer Science and Information Technologies*, vol. 7, no. 3, pp. 1174–1179, 2016.

[11] S. W. A. Shah, H. Ahmad, and M. Rahman, “Performance evaluation of supervised machine learning models for breast cancer diagnosis,” *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021.

[12] Y. Jiang et al., “Artificial intelligence-assisted diagnosis of breast cancer using digital mammography: a review,” *Journal of Cancer Research and Clinical Oncology*, vol. 148, no. 4, pp. 905–919, 2022.

[13] K. P. Singh and A. Kumar, “A robust ensemble machine learning framework for early-stage cancer detection,” *Pattern Recognition Letters*, vol. 164, pp. 45–52, 2024.

[14] R. Subramanian, S. Ahuja, and P. Gupta, “Explainable AI in healthcare: enhancing interpretability of breast cancer models using SHAP and LIME,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 2, pp. 900–912, 2024.

[15] J. Xu, Y. Zhang, and P. Li, “Comparative analysis of feature selection and classification algorithms for breast cancer prediction,” *Applied Soft Computing*, vol. 128, p. 109598, 2022.

[16] K. Polat and S. Güneş, “Breast cancer diagnosis using least-squares SVM,” *Measurement*, vol. 42, pp. 1102–1108, 2009.

[17] S. Bellaachia and E. Guven, “Predicting breast cancer survivability using data mining techniques,” *IEEE ICTAI*, 2006.

[18] S. Chaurasia and S. Pal, “A comparative study of ML algorithms for breast cancer prediction,” *IJCSIT*, vol. 5, pp. 3957–3960, 2014.

[19] S. D. Krishnamoorthi et al., “Deep CNNs for breast histopathology,” *Computers in Biology and Medicine*, vol. 132, p. 104320, 2021.

[20] R. Kumar and A. Singh, “Hybrid genetic algorithm–SVM framework,” *Expert Systems*, vol. 38, no. 1, 2021.

[21] H. Sharma et al., “Breast cancer classification using transfer learning,” *Medical Imaging*, Springer, 2020.

[22] X. Yang et al., “Imbalance handling in cancer classification,” *Information Sciences*, vol. 590, pp. 119–133, 2022.

[23] A. Ahuja and P. Singh, “ML-based radiomics for breast cancer,” *Medical Physics*, vol. 50, pp. 1201–1215, 2023.

[24] G. Litjens et al., “Deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[25] D. Shen et al., “Deep learning in medical imaging,” *Annual Review of Biomedical Engineering*, vol. 19, pp. 221–248, 2017.

[26] A. S. Ashraf et al., “Breast cancer risk prediction using ML,” *Journal of Healthcare Engineering*, 2021.

[27] C. Wahab et al., “A novel ML pipeline for biomedical classification,” *IEEE Access*, vol. 9, pp. 112340–112350, 2021.

[28] R. Jain and R. Chauhan, “Feature engineering in cancer prediction,” *Biocybernetics and Biomedical Engineering*, vol. 41, pp. 1044–1057, 2021.

[29] S. S. Mohanty et al., “Evaluation of ML classifiers for medical data,” *Health Information Science*, 2022.

[30] M. Irvin et al., “CheXpert: a large dataset for medical ML,” *AAAI*, 2019.

[31] R. R. Selvaraju et al., “Grad-CAM for visual explanations,” *ICCV*, 2017.

[32] M. Hussain et al., “Breast cancer prognosis using ML,” *Healthcare Analytics*, vol. 2, 2022.

[33] Z. Qiu et al., “Benchmarking ML algorithms under standardized pipelines,” *Knowledge-Based Systems*, vol. 262, 2023.

[34] M. T. Rahman et al., “Automated diagnosis using hybrid CNN-SVM,” *IEEE Access*, vol. 8, pp. 114142–114154, 2020.

[35] R. Das and A. Turkoglu, “Diagnosis of breast cancer using ML,” *Expert Systems with Applications*, vol. 36, pp. 9821-9825, 2009.

[36] Radiopaedia Spiculated breast cancer(https://radiopaedia.org/cases/spiculated-breast-cancer)

[37] A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45(4):427–437. Doi: 10.1016/j.ipm.2009.03.002

[38] An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861–874. Doi: 10.1016/j.patrec.2005.10.010.

[39] Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J Mach Learn Technol*. 2011;2(1):37–63.

[40] Asri H, Mousannif H, Moatassime HA, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Comput Sci*. 2016; 83:1064–1069.

[41] Cover T, Hart P. Nearest Neighbor pattern classification. *IEEE Trans Inf Theory*. 1967;13(1):21–27. doi:10.1109/TIT.1967.1053964.

[42] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst Appl*. 2009;36(2):3240–3247.

[43] Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297. doi:10.1007/BF00994018.

[44] Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification in breast cancer diagnosis. *Bioinformatics*. 2000;16(10):906–914.

[45] Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. doi:10.1023/A:1010933404324.

[46] Liu Y, Wang Y, Zhang J. New machine learning algorithm: Random Forest. In: *Information Computing and Applications*. Springer; 2012. p. 246–252.

[47] Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001;29(5):1189–1232. doi:10.1214/aos/1013203451.

[48] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016. p. 785–794.