<u>Breast Tumor type detection & prediction using Machine Learning</u>

Project Report submitted in partial fulfillment of
The requirements for the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Of

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY

By

Jayasree Bhattacharya, Roll no: 10900120005
Manan Mal, Roll no: 10900120136
Sarthak Das, Roll no: 10900120156
Subham Pal, Roll no: 10900120141

Under the guidance of

Prof. Rajkumar Patra

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NETAJI SUBHASH ENGINEERING COLLEGE
TECHNO CITY, GARIA, KOLKATA – 700152

<u>ACADEMIC YEAR (2023-24)</u>

CERTIFICATE

This is to certify that this project report titled "Breast Tumor type detection & prediction using Machine Learning" submitted in partial fulfillment of requirements for award of the degree Bachelor of Technology (B.Tech) in Computer Science and Engineering of Maulana Abul Kalam Azad University of Technology is a faithful record of the original work carried out by,

Jayasree Bhattacharya, Roll no: 10900120005, Regd. No. 201090100110168 of 2020-21

Manan Mal, Roll no: 10900120136, Regd. No. 201090100110037 of 2020-21

Sarthak Das, Roll no: 10900120156, Regd. No. 201090100110017 of 2020-21

Subham Pal, Roll no: 10900120141, Regd. No. 201090100110032 of 2020-21

under my guidance and supervision. It is further certified that it contains no material, which to a substantial extent has been submitted for the award of any degree/diploma in any institute or has been published in any form, except the assistance drawn from other sources, for which due acknowledgement has been made.

Date: ___________

_______________	_______________
Dr. Piyali Chatterjee	Prof. Rajkumar Patra
Professor and HOD	Assistant Professor
Department of CSE	Department of CSE
NETAJI SUBHASH ENGINEERING COLLEGE	NETAJI SUBHASH ENGINEERING COLLEGE
TECHNO CITY, GARIA	TECHNO CITY, GARIA
KOLKATA – 700 152	KOLKATA – 700 152

DECLARATION

We hereby declare that this project report titled "Breast Tumor type detection & prediction using Machine Learning" is our own original work carried out as an undergraduate student in Netaji Subhash Engineering College except to the extent that assistance from other sources are duly acknowledged. All sources used for this project report have been fully and properly cited. It contains no material which to a substantial extent has been submitted for the award of any degree/diploma in any institute or has been published in any form, except where due acknowledgement is made.

Student Name:	**Student Signatures:**	**Dates:**
1.	....	....
2.	....	....
3.	....	....
4.	....	....

CERTIFICATE OF APPROVAL

We hereby approve this dissertation titled **"Breast Tumor type detection & prediction using Machine Learning",** carried out by,

Submitted by,

Jayasree Bhattacharya, Roll no: 10900120005, Regd. No. 201090100110168 of 2020-21

Manan Mal, Roll no: 10900120136, Regd. No. 201090100110037 of 2020-21

Sarthak Das, Roll no: 10900120156, Regd. No. 201090100110017 of 2020-21

Subham Pal, Roll no: 10900120141, Regd. No. 201090100110032 of 2020-21

under the guidance of Mr. Rajkumar Patra of Netaji Subhash Engineering College, Kolkata in partial fulfillment of requirements for award of the degree Bachelor of Technology (B. Tech) in Computer Science and Engineering of Maulana Abul Kalam Azad University of Technology.

Date: ___________

Examiners' Signatures:

1. ....
2. ....
3. ....

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to all those who have contributed to the successful completion of our final year project and the preparation of this report. First and foremost, I would like to extend my sincere appreciation to my project guide, Prof. Rajkumar Patra, for his invaluable guidance, support, and expertise throughout this project. His constant encouragement, insightful feedback, and constructive criticism have been instrumental in shaping this work and pushing us to achieve the best results. We are truly grateful for his unwavering commitment and mentorship.

We are also indebted to the faculty members of the Department of Computer Science at Netaji Subhash Engineering College, for their remarkable teachings, which have equipped us with the necessary knowledge and skills to undertake this project. Their passion for the subject and dedication to imparting knowledge have been truly inspiring. Furthermore, we would like to extend our thanks to our friends and classmates who have supported us in various ways, from brainstorming ideas to providing technical assistance. Their collaborative spirit and enthusiasm have made this journey more enjoyable and rewarding.

We would like to express our gratitude to the staff of theNetaji Subhash Engineering College, library and other relevant resources for their assistance in accessing research materials, books, and journals. Their promptness and willingness to help have been crucial in expanding our understanding of the subject matter.

Lastly, We would like to thank our beloved family members for their constant love, encouragement, and belief in our abilities. Their unwavering support and sacrifices have been the pillar of strength that has propelled us forward during challenging times.
We are deeply grateful to all those mentioned above, as well as anyone else who has directly or indirectly contributed to the successful completion of this project. Your invaluable support and guidance have played an integral role in shaping my academic journey, and we will forever cherish the lessons learned and experiences gained throughout this process.

Thank you all.

Jayasree Bhattacharya
Manan Mal
Sarthak Das
Subham Pal

Date: _____________

ABSTRACT

Cancer-related mortality remains a significant challenge, particularly in developing regions where preventive measures are limited or ineffective for certain cancer types. Despite various preventive strategies, some cancers remain unpredictable, complicating clinicians' ability to develop tailored treatment plans. Various algorithms like, K-Nearest Neighbor (KNN), Logistic Regression, support vector machine(SVM) and ensemble learning have been used for predicting breast cancer outcomes using diverse datasets. Our research project aims to select the most suitable predictive approach based on specific needs, contributing to improved accuracy in breast cancer prognosis. Our findings indicate that accuracy of prediction for the XGBoost algorithm increases with a change in training-testing-split ratio of the dataset. The project underscores the significance of precise cancer prognosis in reducing mortality, especially in resource-limited settings. Future research could expand to predict additional variables, further categorizing breast cancer studies and exploring correlations between predictive variables, thereby enhancing our understanding of cancer characteristics and progression. The use of simulation environments and advanced platforms bolsters the reliability of these experiments, potentially leading to broader applications in cancer prediction, detection, and analysis. This study thus contributes significantly to efforts aimed at enhancing the precision of breast cancer prediction and management, offering a structured approach to understanding and forecasting different aspects of the disease.

Table of Contents

Chapter 1: Introduction

Breast Cancer: A malignant cancer that has developed from cells in the breast is referred to as "breast cancer." Breast cancer usually starts in the cells of the milk-producing glands, known as lobules or ducts that drain milk from the lobules to the nipple. Breast cancer can also develop in the breast's stromal tissues, encompassing fatty and fibrous connective tissues.

Prediction: The intention of this study is to create a prediction system that can foretell the onset of breast cancer at an early stage by analyzing the smallest set of attributes from a clinical dataset. The proposed experiment was done utilizing the Wisconsin breast cancer dataset (WBCD).

Analysis and MRI: Breast MRIs are most commonly used in women who have been diagnosed with breast cancer. To assist in measuring tumor volume, searching for further cancers in the breast, and looking for tumor cells in the opposite breast. For those women at high risk of breast cancer, a screening MRI, in addition to a yearly mammogram, is indicated.

Tumor: Breast cancer is a type of cancer that develops in the breast cells. Breast cancer is the second most common cancer diagnosed in women in the United States, after skin cancer. Breast cancer can strike both men and women, but it affects women far more frequently.

Diagnosis: The cells are categorized based on how they appear under a microscope. Breast cancer is classified into two types: invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS), the latter of which advances slowly and has little influence on patients' daily life.

Breast cancer can be effectively treated if detected early. As a result, having access to appropriate means of screening is critical for detecting the first signs of breast cancer. To screen for this disease, many imaging modalities are performed with mammography, ultrasound, and thermography. Mammography is one of the most important early detection methods for breast cancer. Since mammography is inadequate for solid breasts, ultrasound or diagnostic sonography methods are widely used. Radiations from radiography can avoid small masses, and thermography may be more effective than ultrasonography in identifying smaller malignant masses considering these issues. The term "breast cancer" refers to a big tumor formed from breast cells. Breast cancer typically begins in the cells of the lobules, or milk-producing glands, or the ducts, which are passageways that drain milk from the lobules to the nipple. Breast cancer can also start in the stromal tissues of the breast, which include functional areas that are rich and rigid. Cancer cells have the ability to penetrate nearby healthy breast tissue and go to the underarm lymph nodes, which are microscopic structures in the body that filter out alien pathogens.

For so many years there was only one option to detect breast cancer and that is an X-Ray from which medicos used to check the current situation of the patient suffering from breast cancer and then give treatment to them. Now as time has passed many emerging technologies have taken part in medical science and solved many of our quandaries through their technical solutions like MRI, CT scan,etc. The proposed work can be used to predict the outcome of different techniques and suitable techniques can be used depending upon requirement. This work is carried out to predict, detect and analyze the accuracy of breast cancer. Future research can be done to predict other variables, and breast cancer research can be divided into categories based on these variables. Based on the findings of this study, it is reasonable to state that if a patient has a breast cancer tumor, it can be identified.

Chapter 2: Literature Survey

Sl No.	AUTHOR'S NAME	WORK
1.	Larie S, Abukmeil S	By utilizing Ultra Wide Band (UWB) Zhang, introduced a breast tumor detector that uses gelatin-oil technology to obtain experimental results and, as a result, to demonstrate the efficacy of microwave images in the prostate cancer diagnosis.
2.	D. Carvalho	To detect breast cancer diagnosis in an expeditious manner the author proposed a Stochastic pulse synthesizer in UWB application that requires utilizing a static inverter with phase detector to expedite the system detection process.
3.	M K Agrawal, Meena Keshav Kumari	By utilizing odds-ratio curves the author has introduced a logistic Generalized Additive Model (GAM) along with linear Kernel smoother.
4.	R S Barman	The authors utilized the Back propagation Neural Network to detect tumors who have reached the age of 40. Additionally compared the results to another model that used the circular sector function network.
5.	Ikeda, Debra M.	For breast cancer presage the authors have utilized direct subtraction beam composing imager, utilizing numerical simulation and electromagnetism in their model and discovered great resolution and robustness in diagnosing breast cancer.
6.	Clough G R, Truscott J & Haigh I.	The authors have suggested malignant lump must be excised once a patient has been diagnosed with breast cancer. Physicians must determine the disease's prognosis during this procedure. This is a forecast of the disease's expected progression. The term "analysis of survival or lifetime data" is also used to describe the prognosis problem. Since the data is censored, it presents a more difficult problem than diagnosis. That is, there are only a few cases where a disease recurrence has been observed. We can classify the patient as recurrent in this case, and we know when they will recur (TTR). On the other hand, most patients do not experience recurrence.
7.	Alghodhaifi, H., Alghodhaifi, A., Alghodhaifi, and M	The author has made use of Histology analysis to allow for the distinction of benign and malignant tissue. Furthermore, this analysis aids in the performance of a prognostic evaluation. Changes in normal breast tissue structures are referred to as benign lesions. Invasive ductal carcinoma is one type of benign lesion. Algorithms for machine learning (DL) are being introduced to the mainstream audience in order to tackle a variety of image analysis and image recognition difficulties.

8.	Silva, J., Lezama, O.B.P., Varela, N., Borrero	These algorithms were used by the authors, and they produced good classification results, encouraging many researchers to use them to solve difficult problems. In breast histology pictures, a convolutional neural network (CNN) was employed to predict and categories invasive ductal cancer with a granularity of nearly 88 percent. Aside from that, data mining is commonly used in the medical field to predict and classify abnormal events in order to gain improved knowledge of serious diseases like cancer. The findings of employing text mining to detect and classify breast cancer are promising.
9.	B. AK Bugday	In the author's stage of breast cancer (I–IV) indicates cancer spread level in a patient. Stages are determined using statistical indicators such as tumor size, lymph node metastasis, and distant metastasis, among others. Patients must undergo breast cancer surgery, chemotherapy, radiotherapy, and endocrine therapy to prevent cancer from spreading. The researcher's goal is to identify and classify malignant and benign patients, as well as to figure out how to parameterize our classification techniques for high accuracy. We are looking into many datasets and how further Machine Learning algorithms can be used to characterize Breast Cancer.
10.	Abien Fred M. Agarap	The author suggests breast cancer instances are classified as either cancerous or innocuous. The objective of this research is to determine the number of hidden layers, the number of neurons in each hidden layer, and the kind of activation functions in hidden levels in order to build an Artificial Neural Network (ANN) with a high degree of accuracy. The Wisconsin Breast Cancer Database (WBCD), an open-access dataset, provided samples for ANN training and test. There are 699 samples in the dataset, which were split into two groups: The trained model has 599 samples, whereas the test data contains 100 samples. As inputs to the network, each sample has nine attributes that represent nine characteristics of breast fine-needle aspirates (FNAs). When three transfer functions were used, the mean square error (MSE) achieved was equated in this experiment.
11.	Mumine Keles	The author has utilized the most common indicator of breast cancer as the growth of a painless, firm mass in the breast. In the absence of a mass, however, around 10% of patients experience pain. Breast cancer, like other forms of cancer, can be treated more successfully if caught early. Early identification of melanoma improves not only the no of options, but also the likelihood of treatment success and survival.
12.	Viswanatha V., Ramchandra A. C., Avinash Bhagat,	The author has utilized Logistic Regression, a binary classification algorithm which is suitable to solve binary classification problems like breast cancer detection.

	Shashank Shekhar	Logistic Regression predicts binary outcomes using a sigmoid function converting inputs to probabilities by leveraging input features like tumor size, age and other relevant data, logistic regression models predict the likelihood of the presence of breast cancer. This predictive capability makes it a valuable tool for early diagnosis allowing healthcare professionals to identify potential cases and initiate timely interventions. The simplicity and interpretability of Logistic Regression contribute to the effectiveness of early detection and management of breast cancer.
13.	Prof. Alok Chauhan, Harshawardhan Kharpate, Yogesh Narekar, Sakshi Gulhane, Tanvi Virulkar, Yamini Hedau	The author has used MRI images of the patient to extract features and detect cancer or the tumor at the exact position in the MRI by using algorithms including SVM, KNN and CNN. SVM has been used to classify breast cancer into malignant or benign. KNN has been used to solve both relegation and regression problems which is helpful to comprehend the results. This technique employs a neural network to detect breast cancer in scans of breast tissue. The proposed methodology of our paper is based on this [13].
14.	Rahmanul Hoque, Suman Das, Mahmudul Hoque and Ehteshamul Haque	The author has used the XGBoost model and has tried to improve its accuracy. He has compared it with other machine learning models including SVM, KNN, Logistic Regression etc.

Chapter 3: Proposed Work

The proposed system will use machine learning algorithms and data visualization methods for accuracy. The system will first collect related data such as medical history, family history, and mammography images as shown in Fig 1, and preprocessing data to make sure they are clean and structured. To find the most suitable features of breast cancer forecasting, the system will use various methods of feature selection and function engineering. Machine Learning (ML) algorithms such as logistic regression, ensemble learning techniques and K-NN will be used to train the classification model using selected properties. The system will use data visualization methods to visualize the most important aspects and provide insights into the decision-making process, increasing the interpretation of the model. The system will also use explained artificial intelligence methods to increase the model's openness and provide reasons for its predictions. Finally, the system will be rated from a huge collection of mammography photos to assess its accuracy and performance. A proposed flowchart is shown in Fig 2 which describes our model and the steps mentioned in the flowchart are discussed below.

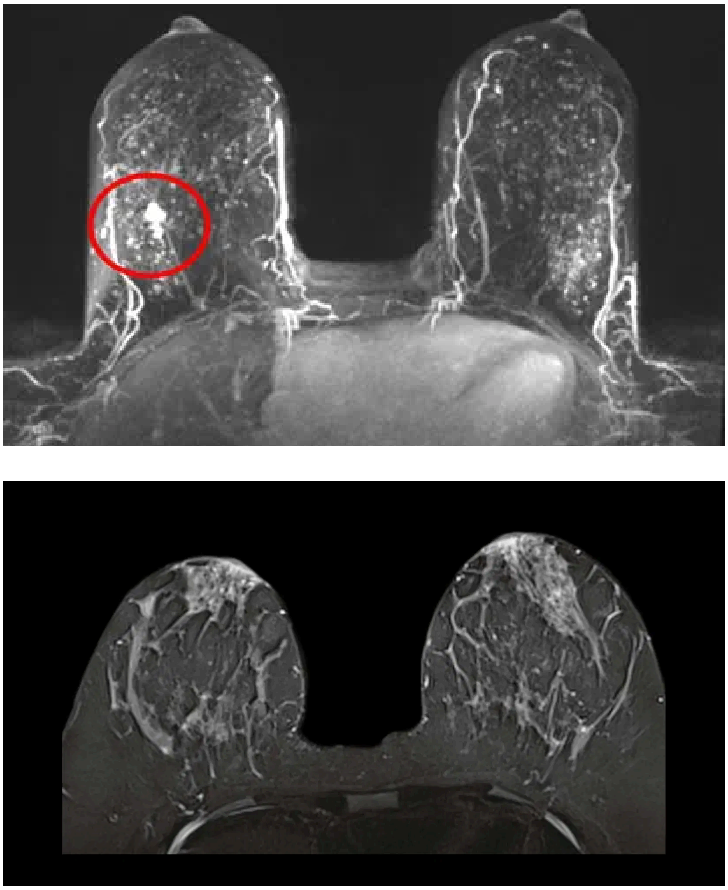

Fig 1. X-Ray of Human Breasts

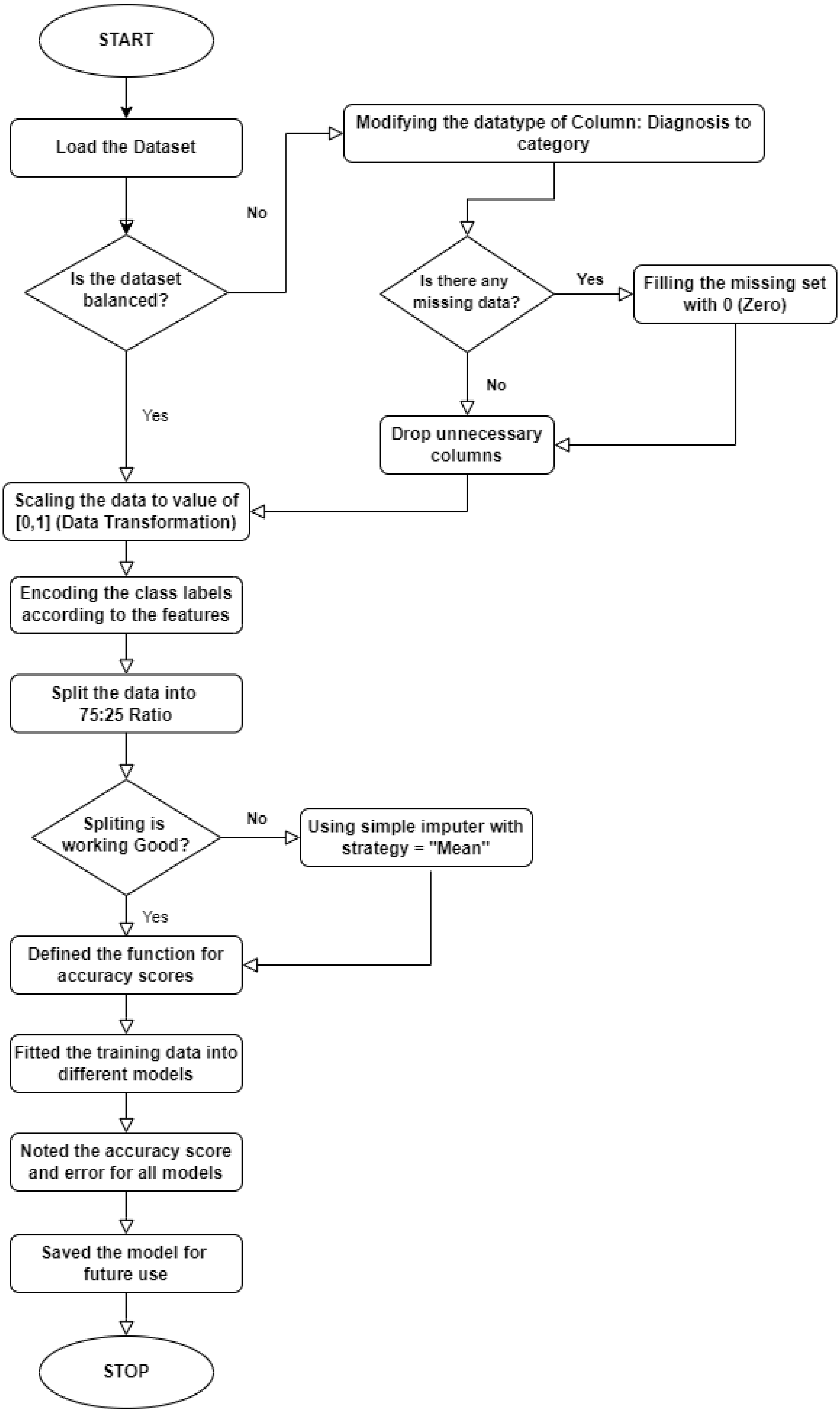

Fig 2. Proposed flowchart of model

3.1 Dimensionality reduction : By removing the less important factors, The total amount of independent variables is decreased to a few key factors. crucial in determining the result. Dimensionality reduction is used to transform three-dimensional data into two dimensional data so that the machine learning (ML) models may be more successfully visualized by charting the predicted areas and bounds for every model. Regardless of the number of independent variables, adopting a good dimensionality reduction approach typically results in two independent variables. There are two techniques for selecting features: “feature selection and feature extraction”.

3.2 Data Loading and Preprocessing : The project starts by importing necessary libraries such as NumPy and pandas. The scikit-learn library is used to load the Wisconsin Diagnostic Breast Cancer dataset. The dataset contains measurements from fine needle aspirates of breast masses and is split into input features (X) and the target variable (Y). After loading the dataset, it is converted into pandas DataFrame (data_frame), which facilitates data manipulation and analysis. The DataFrame includes the input features as columns and the target variable as the 'label' column.

3.3 Feature selection and projection : Building an accurate and effective breast cancer prediction model requires careful feature selection. Finding the key characteristics that are most crucial for predicting breast cancer outcomes is the aim of feature selection. The proposed system will find the most crucial features for breast cancer prediction using a variety of feature projection techniques. With feature projection as shown in Fig 3, the data’s dimensionality is decreased while the most crucial features are kept. The machine learning model’s performance can be enhanced by this process, which can also lessen overfitting. Because of the kind of feature in the dataset, we can employ both linear and non-linear reduction. The dataset used in the research is multidimensional. Moreover, it can not give the most accurate features. So that’s why we applied the feature projection technique.

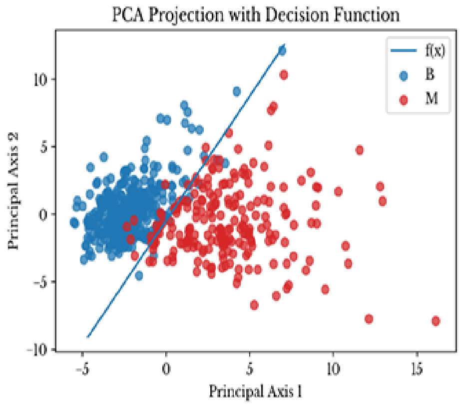

Fig 3. PCA Projection

3.4 Data Exploration and Visualization : Exploratory data analysis is performed to gain insights into the dataset. The shape of the dataframe is checked to determine the number of rows and columns. Information about the dataset, including data types and missing values, is obtained and descriptive statistics, such as mean, standard deviation, and quartiles, are calculated. The features are then transformed into value [0,1] To visualize the relationship between the features and the target variable, scatter plots are created as shown Fig 4. Each feature is plotted against the target variable, with different marker colors representing benign or malignant tumors. These visualizations help in understanding the distribution of feature values and potential separability between the two classes.

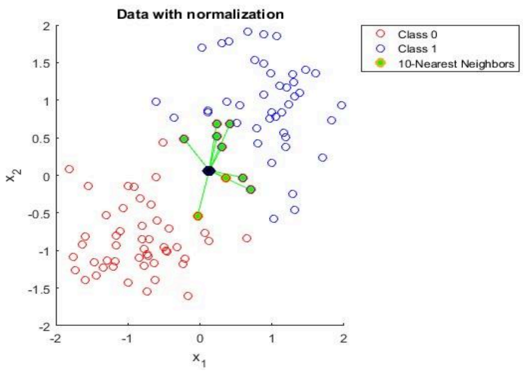

Fig 4. Normalization Graph

3.5. Model Training : Model selection is the practice of deciding which predictive model is most appropriate to use for a given dataset when making predictions using data visualization and machine learning. In this procedure, various machine learning (ML) algorithms are applied to a dataset, and the effectiveness of each algorithm is assessed using a variety of metrics like recall, accuracy, and precision. Visualization helps us to discover which features are most essential for forecasting the outcome of interest, which can inform model selection. Once the best-performing model has been determined, It might be used to create predictions based on fresh data. We trained these models on a collection of breast tumor data, which included characteristics including tumor size, shape, and texture. Following training, we assessed each model's performance and compared its accuracy in predicting malignant and benign tumors. The different types of classification algorithms that we used are the following,

3.5.1 Logistic Regression : This is a binary classification algorithm and a supervised learning classification algorithm which is used to predict probability of a target variable. Here, the nature of the target or dependent variable is dichotomous, which means there will be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/Yes) or 0 (stands for failure/No).

Mathematically, a LR model predicts P(Y=1) as a function of X. It is one the simplest machine learning algorithms that can be used for classification problems such as breast cancer detection and prediction. A logistic regression model is instantiated using the LogisticRegression() class from scikit-learn. The model is then trained on the training data using the fit function, which estimates the optimal parameters that maximize the likelihood of the observed data.

3.5.2 K-nearest neighbor (k-NN) : This is a pattern recognition technique that uses training datasets to find k's closest relatives in subsequent samples. The theory of the adjacent method is used to describe a set of training datasets near the new point, which are then used to anticipate the breakpoint. As in k-nearest neighbor (k-NN) learning, the user can choose the specimen or change it based on the local point density.

3.5.3 Bagging (Bootstrap Aggregating) : In bagging, multiple instances of the same learning algorithm are trained independently on different subsets of the training data. These subsets are often created by random sampling with replacement (bootstrap sampling). The predictions of each model are then combined, usually by averaging (for regression) or voting (for classification), to produce the final ensemble prediction.

3.5.4 Boosting : Boosting involves training a sequence of weak learners sequentially, where each new model focuses on correcting the errors made by the previous ones. During training, instances that were misclassified by earlier models are given higher weights, emphasizing them in subsequent rounds.The final prediction is a weighted sum of the individual weak learners, with more weight assigned to those that performed well on the training data.

While working with these models, we divided the dataset into a 75:25 training-testing split. We imported and trained the model with 75% of the dataset, using the remaining 25% for testing to evaluate its predictions. To assess the model's performance, we imported accuracy_score, classification_report, confusion_matrix, and cohen_kappa_score from sklearn.metrics to determine the accuracy, generate a confusion matrix, and calculate the kappa value.

3.6 Model Evaluation : The above trained models are evaluated using performance metrics. The accuracy score is calculated for both the training and test data using the accuracy_score() function, which compares the predicted labels with the true labels. The accuracy metric indicates the proportion of correctly classified samples.

3.7 Prediction : The input data is converted into a numpy array and reshaped to match the expected format. The trained models then predict the malignancy of the tumor using the predict() function. The prediction output is displayed along with an accompanying message indicating whether the breast cancer is classified as benign or malignant.

Chapter 4: Result and Discussion

4.1 Description of the Dataset : The performance of the machine learning techniques is based on the availability of a suitable and valid dataset. As per the documentation of the dataset, the dataset [20] encompasses attributes pertaining to the attributes of cellular nuclei. These attributes are derived from a digitized representation of a fine needle aspirate (FNA) of a breast mass. The dataset comprises 569 rows, where each row denotes a distinct digitized image of a breast mass, and 33 columns. Out of 569 rows, no column is missing data besides the "Unnamed: 32" column which only has null values. The "Unnamed:32" column along with the "id" column has been removed since neither will be useful for our analysis. Each digitized image either contains benign (i.e. non-cancerous) or malignant (i.e. cancerous) breast cells. We should determine how many are benign and how many are malignant. The total number of malignant (M) and benign (B) samples is 212 and 357, respectively. So, out of 569 instances in our dataset: 357 contained benign breast cells (62.74%), 212 contained malignant breast cells (37.26%).

4.2 Result : We have calculated the accuracy, kappa score, precision and recall for all the models as shown in Table 1. We have constructed an accuracy graph for all models used as shown in Fig 5 and also made a confusion matrix for the proposed model as shown in Fig 6. When using K-Nearest Neighbour, Random Forest and Logistic Regression, we expect a higher success rate of the model such as 95% - 99%. So, if the model gives us the accuracy level in between 95% - 99% then we can conclude that the selection of features is successful according to the dataset.

Table 1: Comparison of other methods with XGBoost

Performance Measurement	XGBoost	Logistic Regression	K-NN (K=5)	Random Forest	Bagging Technique
Classification Accuracy (in %)	97.2027	96.5034	95.8041	96.5034	95.1048
Kappa Score	0.9404	0.9247	0.9107	0.9258	0.8962
Precision	0.9775	0.9433	0.9662	0.9662	0.9550
Recall	0.9775	0.9615	0.9662	0.9772	0.9659

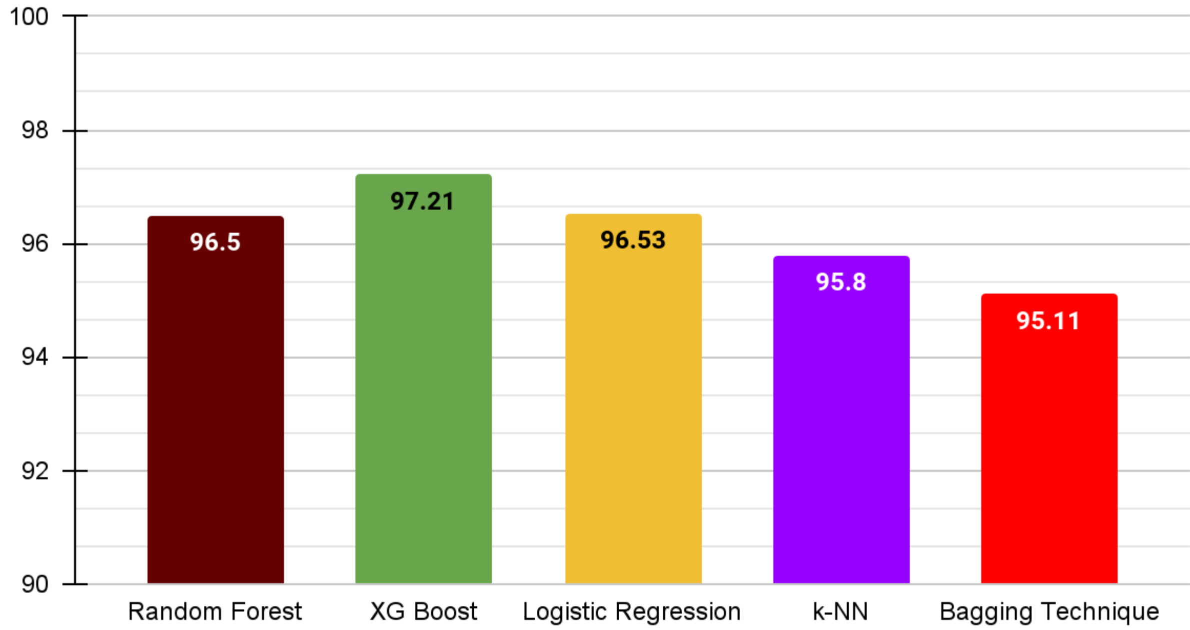

Fig 5: Accuracy Graph

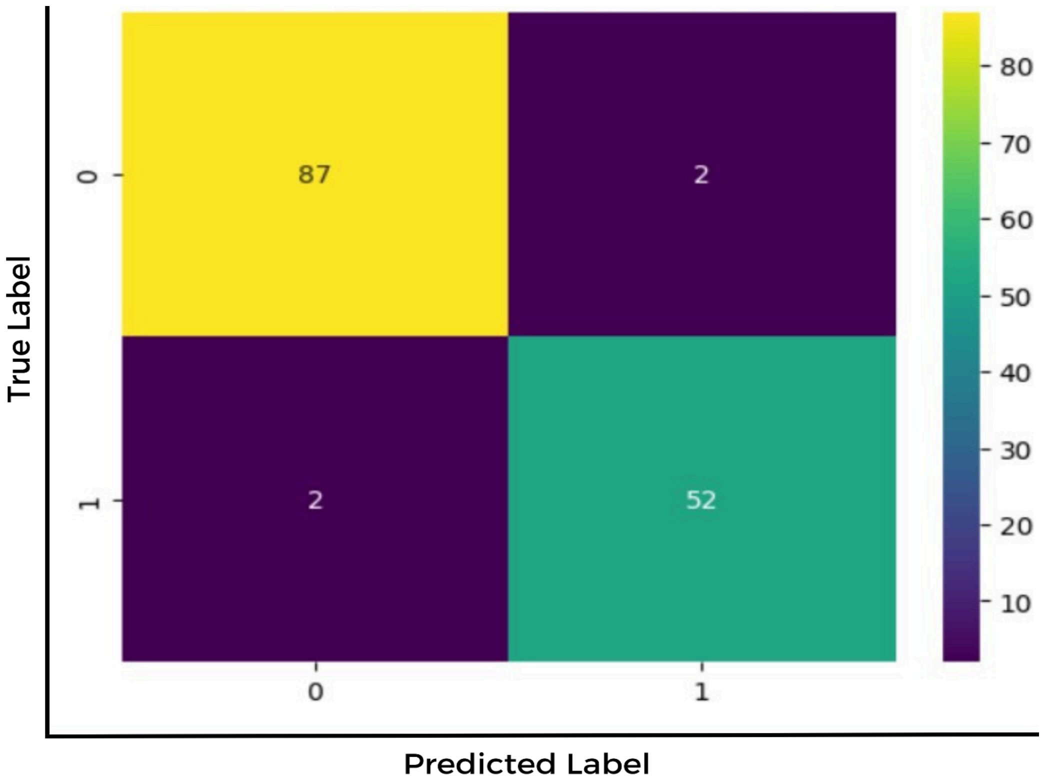

Fig 6: Confusion Matrix for XGBoost

4.3 Result Analysis : We have done the research by splitting the dataset into training and testing data of 25% testing size and 75% training size. From a report presented by [14] Rahmanul Hoque, Suman Das, Mahmudul Hoque and Ehteshamul Haque on Breast Cancer Classification using XGBoost, we compare our result with this previous work as shown in Table 2 and Fig 7.

Table 2: Comparison of our and previous work

Performance measurement of XG Boost	Previous Work Accuracy	Our Work Accuracy
Classification Accuracy	94.74%	97.20%
Precision	0.9444	0.9775
Recall	0.9714	0.9775

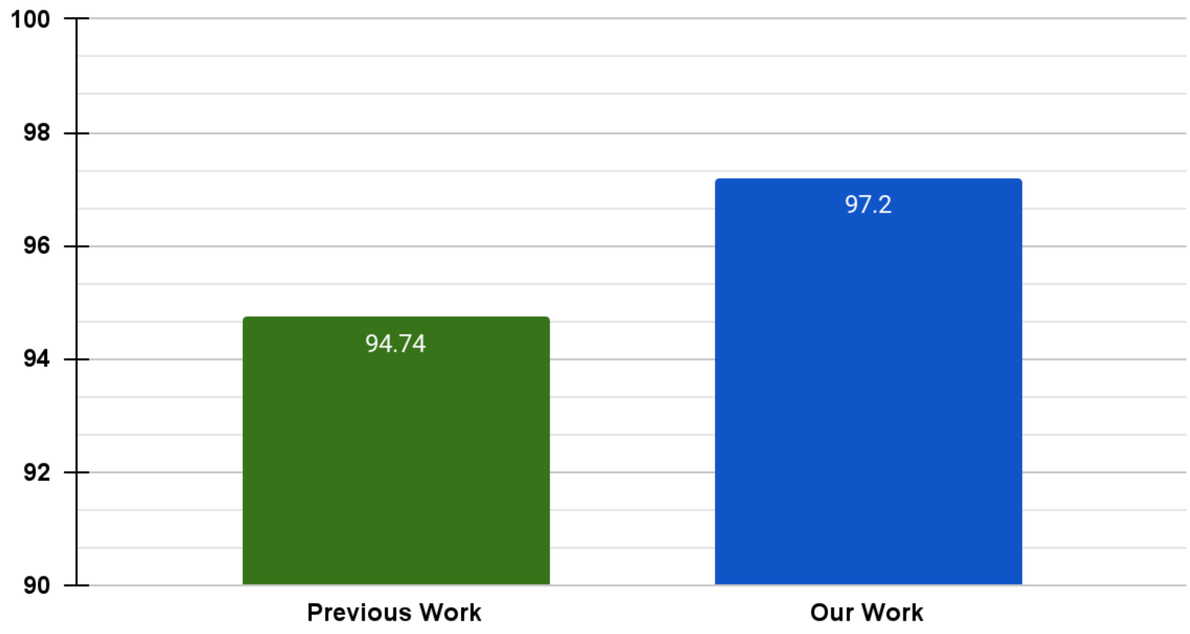

Fig 7: Accuracy Comparison Graph

Chapter 5: Conclusion

The relative performance of these algorithms depends on various factors such as dataset characteristics, feature engineering, data normalization and hyperparameter tuning . It's recommended to perform a thorough evaluation by implementing both logistic regression, KNN and ensemble learning algorithms on the Wisconsin Breast Cancer dataset, using techniques like cross-validation, and comparing their accuracy, precision, recall, and other relevant metrics.

To conclude we have observed that the performance of an algorithm changes depending on the training-testing data split. We have applied this theory on the XGBoost Algorithm and found out that it produces better results as shown in Fig 7. In [14] Rahmanul Hoque, Suman Das, Mahmudul Hoque and Ehteshamul Haque on Breast Cancer Classification using XGBoost, they have utilized 80:20 split ratio while we have tested using 75:25 split ratio and the algorithm showed improved accuracy and precision.

References:

[1]. Larie, S., & Abukmeil, S. (2020). A systematic and quantitative review of magnetic resonance imaging. *Journal of Psychiatry*, 172, 110-120.

[2]. Carvalho, D. (2019). A hybrid model to support the early diagnosis of breast cancer. *Procedia Computer Science*, 91, 927–934.

[3]. Agrawal, M.K., & Kumari, M.K. (2019). Tumor prediction system. *International Journal of Research in Engineering and Technology (IJRET)*.

[4]. Barman, R.S. (2020). Cancer detection using histopathological images. *IEEE*.

[5]. Ikeda, D.M. (2020). *Breast Imaging: The Requisites*. Philadelphia, PA: Elsevier Inc.

[6]. Clough, G.R., Truscott, J., & Haigh, I. (2019). Can high frequency ultrasound predict metastatic lymph nodes in patients with invasive breast cancer? *The Society and College of Radiographers*, 12, 96-104.

[7]. Alghodhaifi, H., Alghodhaifi, A., & Alghodhaifi, M. (2019). Predicting invasive ductal carcinoma in breast histology images using convolutional neural network. In *2019 IEEE National Aerospace and Electronics Conference (NAECON)*, pp. 374–378.

[8]. Silva, J., Lezama, O.B.P., Varela, N., & Borrero, L.A. (2019). Integration of data mining algorithms and ensemble learning for predicting breast cancer recurrence. In Miani, R., Camargos, L., Zarpelão, B., Rosas, E., & Pasquini, R. (Eds.), *Lecture Notes in Computer Science* (Vol. 11484, pp. 18–30). Springer, Cham.

[9]. Bugday, B.A. (2019). Classification of breast cancer data using machine learning algorithms. *Medical Technologies Congress (TIPTEKNO)*, Izmir, Turkey, pp. 1-4.

[10]. Agarap, A.F.M. (2019). On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset.

[11]. Keles, M. (2019). Breast cancer prediction and detection using data mining classification algorithms: A comparative study. *Journal of Computer Applications*, February.

[12]. Viswanatha, V., Ramchandra, A.C., Bhagat, A., & Shekhar, S. Breast cancer classification using logistic regression. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI)*.

[13]. Chauhan, A., Kharpate, H., Narekar, Y., Gulhane, S., Virulkar, T., & Hedau, Y. Breast cancer detection and prediction using machine learning. In *Proceedings of the International Conference on Computational Intelligence and Data Science (ICCIDS)*.

[14]. Hoque, R., Das, S., Hoque, M., & Haque, E. (2024). Breast cancer classification using XGBoost. *World Journal of Advanced Research and Reviews*, 21(02), 1985–1994.