BREAST CANCER DETECTION USINGLOGISTIC REGRESSION

Ch. Srividya[1]*, M. Madhubala[2], Ushakela Neelaveni[3], Abhinav Reddy Kommareddy[4], Dheeraj Davuluri[5]

Abstract—

Best Unregulated chest cell proliferation directly causes cancer. Both women and occasionally people experience it. It is the second most convincing justification for deaths caused by unfavorable development in women worldwide. A woman is likely to die from a harmful chest development 1 out of every 38 times (or 2.6% of the time). Since 2007, the passing rates from chest-threatening development in women over 50 who are still active have been stable, but they have begun to decline in women who are more prepared. Between 2013 and 2017, the mortality rate constantly declined by 1.3 percent. Either the lobules or the chest courses are filled with cells that are in danger of developing. Organs that produce milk are known as lobules, and lines are the routes that transport the milk from the organs to the areola. In this article, we provided a strategy for controlling the assurance of a chest hazardin light of a number of data points that illustrate a few sick cell features. This technique makes use of the AI model known as Logistic Regression. This model is quantifiable and superior accuracy, according to preliminary discoveries.

Keywords— Breast cancer, Machine learning, Logistic Regression, Group method data handling.

[1]*Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- chsrividya12@gmail.com

[2]Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- baladandamudi@gmail.com

[3]Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- ushakelaneelaveni@gmail.com

[4]Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- abhinavkommareddy1@gmail.com

[5]Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- dheerajdavuluri3@gmail.com

*Corresponding Author: - Ch. Srividya
*Computer Science and Engineering, Institute of Aeronautical Engineering,
E-mail:- chsrividya12@gmail.com

I.INTRODUCTION

Breast cancer is a condition in which breast cells grow uncontrollably. There are several types of breast cancer. Thekind relies on which cells in the breast develop into cancer. According to data from the IARC (Indian Astrobiology Research Centre), breast cancer accounts for 25% of allcancer cases diagnosed in women globally. The industrialized nations, which make up 82 percent of theworld's population, account for about 53% of these instances.Invasive breast cancer cases are expected to increase by about 276,480 in 2020. The study of medical imaging now includes machine learning as a key component. Machine learning methodologies have developed over time, moving from manually seeded inputs to the automated initialization we see today. Computer-assisted diagnosis (CAD) systems have become more intelligent and self-sufficient as a result of advancements in machine learning, which are constantly enhancing the learning capacity of these technologies. Based on the patient's attribution, we have suggested the logistic regression approach to determine whether the tumor is malignant or benign. By enabling therapy while the cancer isstill in the early stages of development, early diagnosis greatly improves the outlook for women with breast cancer. Machine learning is a tool that enables computers to automatically learn and grow from information without being expressly designed. Smart methods to studying huge amounts of data are provided by machine learning. By creating rapid, effective algorithms and data-driven models for in-the-moment data processing, machine learning may produce exact findings and analysis. Machine learning is a commonly utilized method in the field of machine imaging, namely in the prediction of a cancer diagnosis.

Logistic regression can be an effective modeling tool for detecting whether a tumor is benign or malignant since it cancompute the chance that a tumor is malignant dependent on a variety of predictor parameters, including the size, shape, and other characteristics of the tumor. Logistic regression may be used to classify tumors by determining which ones are benign and which ones are malignant depending on the features of the tumor. The tumor is categorized as malignant if the predicted likelihood is higher than a certain threshold, which is typically 0.5; otherwise, it is categorized as benign. Overall, logistic regression is a useful tool for binary classification issues, such as evaluating whether a tumor is benign or malignant, as it provides a simple and intuitive approach for estimating class probabilities based on predictor variables. The logistic regression method, a type of extendedlinear model, is used to model the probability of an event occurring as a function of one or more predictor variables. Binary logistic regression's two potential outcomes are frequently denoted by the numbers 0 and 1. The logistic regression model predicts the likelihood that an observation belongs to the positive class (i.e., 1) based on its predictor components.

Logistic regression may be used to calculate the likelihood that a tumor is malignant given its features in the context of determining whether it is benign or malignant. Tumor size, form, location, and other elements that may be important for tumor categorization are examples of these features. Predictions for fresh observations can be generated based on the predictor values once the model has been fitted.The classification of a tumor as benign or malignant depends on whether the estimated chance of a tumor being malignant exceeds a certain threshold, which is typically 0.5.

Random forests are a more sophisticated algorithm than logistic regression, which is quite straightforward. To model the link between the predictor factors and the response variable, random forests employ decision trees. These decision trees have a great deal of adaptability and are capable of capturing intricate nonlinear interactions between the variables.

Logistic regression and random forests have the potential to accurately identify benign from malignant tumors in the context of breast cancer screening. If interpretability is critical or if there is a clear link between the variables, logistic regression could be a viable option. If high accuracyis important to you or if your data has complicated inter- variable relationships, random forests could be a viable option. While training and optimizing random forests, a lot of computing resources may be needed, especially when dealing with huge datasets. On the other hand, logistic regression is often more computationally effective and is capable of doing well even with very little datasets.

The easiest technique to diagnosis breast cancer with the highest degree of accuracy is described in this paper's discussion of breast cancer detection using logistic regression. Based on the characteristics in the provided dataset, it forecasts whether the patient in question has a benign or malignant tumor.

II. LITERATURE REVIEW
The author used real time data set from Wisconsin

breastcancer data. The data set content 569 data points with 30 attributes. Different features of data were mean radius mean talk, show mean perimeter and each of these features described breast lumps. The author has used different algorithms like decision tree algorithm, navies algorithm, support, vector machine, KNN has first plotted a scatterplot of the diagnostic data where it is classified as benign and theoutput into malignant and benign. The classification was based on how the column data of samples were obtained from the coordinates X and Y. They have plotted a logistic curve by calculating the risk score and then implemented the sigmoid function. Yeah, with conclusions their classificationaction accuracy was 90.48% and testing classification was 96% yavudadru results demonstrated both ability of method to diagnose breast cancer and also it is it time saving trainingphase. [Refer section VI-1].

The author has used and simple learning methods for cancer production. The author employed a variety of methods, including KNN, SVM decision trees, logistic regression, and artificial neural networks. gradient descent and naive bayes. The top five features highlighted or glucose resistant HOMA, insulin, BMI and chi-Square. after loading data set, the data set would be analysed and other features would be selected. The model building was done by listed algorithms and dysentery KNN and SVM ended up with heterogenous, assemble staking. Different evolution measures were used in the model study, including AUC, Aro, sick of F1 score, accuracy, recall, log laws, and specificity utilizing Ke fold. The highest level of accuracy was 78%. K complete model evaluation marked the end of the model's optimal performance. [Refer section VI-2]

The author has also included a pinch of neural networks to predict the best breast cancer. There are three existing datasets WBCD , WDBC and WPBD is data sets were used in many of the previous works as yet well but that stated by author. The dataset contains 699 data points 10 features explaining different type of characteristics of the clamp it classified into Benign and malignant. The process included the data sets, splitting into training and testing samples,where each of them are proceeded into minimum maximum normalisation as part of pre-processing. The variables are then sent into feature selection using logistic regression. The test samples or trained GMDH classifier and then classified into Benign and malignant. Whereas the train data set wouldbe training the GMDH classifier and then splitted into two classes. Accuracy, recall precision

ERF measure were calculated on the basis of confusion Matrix. [Refer section VI-3]

The author has demonstrated the use of KNN, logistic regression, and sample learning for the identification of breast cancer. The author used a KNN support vector machine and an artificial neutral network to solve the problem and get highly accurate and precise findings. According to the analysis of confusion Matrix, logistic regression output resulted in 97% of accuracy and testing setgave 99% of accuracy. Hence, all the five machine learning algorithms were compared with their evaluation metrics. The author mentioned that the future scope of this project would be increase helps in better training of the machine, learning algorithm and model for good working more and more accuracy. This would also brief the readers about the relationship between various attributes. [Refer section VI-4]

The author has resplendently proposed Kaizen programming approach with logistic regression, specifying the number of deaths and the reasons of the death by proposing a model with the use of Wisconsin diagnostics breast cancer dataset the other house first derived the BMI from mass and height of a person and later stated that it is biased and time consuming the plan was to propose new ideas into current features, expanding the set of features. Thefeatures were selected and then turned out to be a continuouscycle. The project's novelty lay in the model they proposed for a dataset that was widely used, the high-quality features they identified, and the best results they produced. Keys and programming was a brand-new tool based on the principles of Kaizen methodology that assisted in feature performance, construction, feature selection, and model generation. The existing feature was expanded with the aid of feature creation. Set Kaizen feature selection eliminated critical features, and the algorithm assisted model building. [Refer section VI-5]

We have made an effort to contrast our findings with other methods of cancer detection. The majority of field research focuses on improving patient care through better hospital infrastructure and pharmaceutical administration. Numerous researches have been done on the subject of early cancer diagnosis and treatment, which has the potentialto save lives. Even the few studies that are now available in the field lack implementation details and are cancer-specific, i.e., there are limitations in terms of both geography and wellbeing..

Current system only takes 40% of training dataset which doesn't extract all the features of the patient. It has a low efficiency since lack of features with low accuracy with 76%. Time complexity is high, where it takes more time for compiling. Most cancer cases are mishandled or unresolved as a result of late detection. The fundamental limitation of logistic regression in the detection of breast cancer is the assumption of linearity between the dependent variable and the independent factors. In addition to providing an assessment of a predictor's appropriateness (coefficient size),it also provides information on the correlation's intensity (positive or negative).

In order to determine an individual's cancer stage, our study focuses on creating a breast cancer detection system that uses supervised machine learning algorithm Logistic Regression. By examining the variables of a given dataset, our suggested method can determine if a patient has a benign or malignant tumor. To train and assess our model, we divided the dataset's data into 90% for training and 10% for testing. To gauge the accuracy, recall, and precision of our system, we test our model using a variety of performance measures, such as the confusion matrix, ROC, and AUC. Our suggested approach has the potential to enhance the detection of breast cancer and advance the field of medical research.

III. METHODOLOGY

The methods that are indulged with some basic and complicated progress where the primary dataset is been preprocessed and scaling process takes place at first place followed by the data is splitted into two varies portions namely training dataset and testing dataset with a percentage of 90 and 10 respectively. Then the trained data is sent into feature extraction for extracting the features of the dataset then sent to feature selection and logistic regressionalgorithm is applied and sent for model evaluation, parallelly the testing data is directly sent to the model evaluation for working on testing data with the help of training dataset. Fig.1 determines the block diagram of Breast Cancer Detection using Logistic Regression.
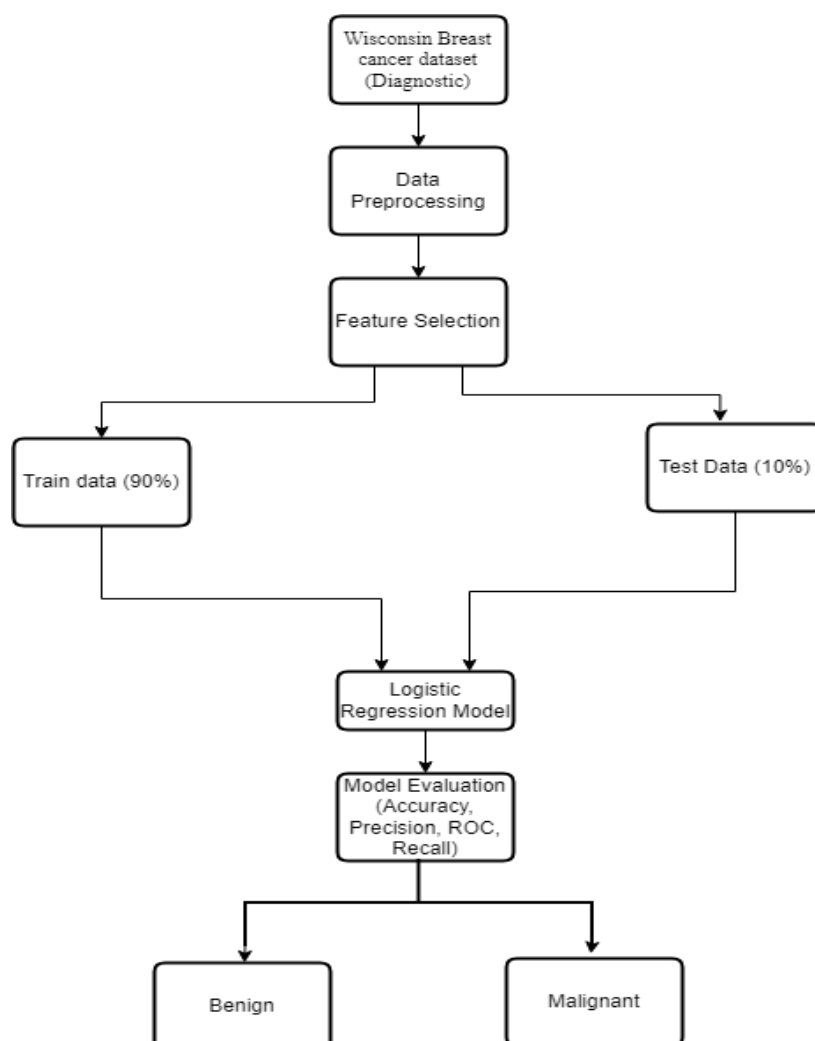

Fig.1 Block Diagram of Breast Cancer Detection using logistic regression

A. Dataset

The Wisconsin Breast Cancer (Diagnostic) dataset provides information about the characteristics of breast masses based on an image of a fine needle aspirate that has been digitally captured. The dataset consists of 569 instances, 32 characteristics, and 10 real-valued features computed for each cell nucleus, including radius, texture, perimeter, area, smoothness, compactness, concavity, concave spots, symmetry, and fractal dimension, as well as ID number and diagnosis (Benign or Malignant). For each image, the mean, standard deviation, and "worst" or greatest of these traits were calculated, yielding a total of 30 features. All feature values are recorded with four significant digits, and the dataset is clean with no missing values. There are 357 benign and 212 malignant students in the class. The purpose of the study is to divide the data into these two categories using logistic regression, and to assess performance using several measures including precision-recall, accuracy, ROC, and confusion matrix.

B. Data Preprocessing

The dataset underwent preprocessing in which the Standard Scaler package was used to normalize the characteristics. Many machines learning algorithms, including Logistic Regression, demand that all features be on the same scale and have a normal distribution. This transformation is done to meet these requirements. Each feature is scaled by the Standard Scaler library to have a mean of 0 and a standard deviation of 1. This preprocessing step enhances the algorithm's performance and speeds up its convergence during training.

C. Feature Selection

There are a total of ten characteristics that are utilized to determine if breast cancer is benign or aggressive. These characteristics include measurements of the radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. To determine the most pertinent characteristics and minimize the amount of data dimensions, feature selection is crucial. This can enhance the model's functionality and lessen the chance of overfitting. Based on the observed feature values, a model may be trained to predict whether breast cancer is benign or malignant using the specified characteristics.

D. Logistic Regression

The commonly employed method for logistic regression Unsupervised machine learning methods for categorization include logistic regression. It is a statistical model that, in its simplest version, models a binary dependent variable usinga logistic function. An observation's likelihood of falling into a certain class or classification is anticipated. The paradigm of linear regression is transformed by logistic regression into Lasso regularization techniques are the most popular. These two well-liked techniques avoid overfitting. By punishing the cost function, the regularization approach is utilized to address the overfitting difficulties. L1 or Lasso regularization and L2 or Ridge regularization are the two regularization methods utilized for processing.

A linear regression technique can only provide binary values of 0 or

1. Therefore, it may be expressed as true or untrue, orprecisely in a this or that pattern, as yes or no.

Algorithm:

import from sklearn.linear_model Logistic Regression
classifier=Logistic Regression() classifier.fit (x_train,y_train)
using sklearn.metrics Import accuracy_score prediction_on_training_data
accuracy_on_training_data = accuracy_score(y_train,prediction_on_training_data);
"Accuracy on training data:" is printed after accuracy_on_training_data.
Confusion matrix is defined by
[[TN FP]
[FN TP]]

E. Model Evaluation

We may assess the breast cancer diagnostic prediction model using measures like accuracy, precision, recall, and F1 score. The Proportion of accuracy properly diagnosed malignant cases among all anticipated malignant cases, whereas accuracy measures the percentage of correctly classified occurrences. Recall quantifies the percentage of real malignant cases that were accurately categorized as such. When the class distribution is unbalanced, the F1 score, which is the harmonic mean of accuracy and recall, is an effective statistic to utilize.

The four main elements of a confusion matrix for the identification of breast cancer are typically:

True Positive (TP): A malignant case is appropriately identified as such by the model.
True Negative (TN): A benign instance is appropriately identified as benign by the model.
False Positive (FP): A Type I mistake occurs when

the model expects a benign instance to be cancerous.

False Negative (FN): The model forecasts a cancer case as benign when it is actually benign (Type II

mistake).

IV. RESULTS AND DISCUSSION

mpactness error	concavity error	concave points error	symmetry error	fractal dimension error	worst radius	worst texture	worst perimeter	worst area	worst smoothness	worst compactness	worst concavity	worst concave points	worst symmetry	worst fractal dimension	class
569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
0.025478	0.031894	0.011796	0.020542	0.003795	16.269190	25.677223	107.261213	880.583128	0.132369	0.254265	0.272188	0.114606	0.290076	0.083946	0.627417
0.017908	0.030186	0.006170	0.008266	0.002646	4.833242	6.146258	33.602542	569.356993	0.022832	0.157336	0.208624	0.065732	0.061867	0.018061	0.483918
0.002252	0.000000	0.000000	0.007882	0.000895	7.930000	12.020000	50.410000	185.200000	0.071170	0.027290	0.000000	0.000000	0.156500	0.055040	0.000000
0.013080	0.015090	0.007638	0.015160	0.002248	13.010000	21.080000	84.110000	515.300000	0.116600	0.147200	0.114500	0.064930	0.250400	0.071460	0.000000
0.020450	0.025890	0.010930	0.018730	0.003187	14.970000	25.410000	97.660000	686.500000	0.131300	0.211900	0.226700	0.099930	0.282200	0.080040	1.000000
0.032450	0.042050	0.014710	0.023480	0.004558	18.790000	29.720000	125.400000	1084.000000	0.146000	0.339100	0.382900	0.161400	0.317900	0.092080	1.000000
0.135400	0.396000	0.052790	0.078950	0.029840	36.040000	49.540000	251.200000	4254.000000	0.222600	1.058000	1.252000	0.291000	0.663800	0.207500	1.000000

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry	mean fractal dimension	radius error	texture error	perimeter error	area error	smoothness error
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	14.127292	19.289649	91.969033	654.889104	0.096360	0.104341	0.088799	0.048919	0.181162	0.062798	0.405172	1.216853	2.866059	40.337079	0.007041
std	3.524049	4.301036	24.298981	351.914129	0.014064	0.052813	0.079720	0.038803	0.027414	0.007060	0.277313	0.551648	2.021855	45.491006	0.003003
min	6.981000	9.710000	43.790000	143.500000	0.052630	0.019380	0.000000	0.000000	0.106000	0.049960	0.111500	0.360200	0.757000	6.802000	0.001713
25%	11.700000	16.170000	75.170000	420.300000	0.086370	0.064920	0.029560	0.020310	0.161900	0.057700	0.232400	0.833900	1.606000	17.850000	0.005169
50%	13.370000	18.840000	86.240000	551.100000	0.095870	0.092630	0.061540	0.033500	0.179200	0.061540	0.324200	1.108000	2.287000	24.530000	0.006380
75%	15.780000	21.800000	104.100000	782.700000	0.105300	0.130400	0.130700	0.074000	0.195700	0.066120	0.478900	1.474000	3.357000	45.190000	0.008146
max	28.110000	39.280000	188.500000	2501.000000	0.163400	0.345400	0.426800	0.201200	0.304000	0.097440	2.873000	4.885000	21.980000	542.200000	0.031130

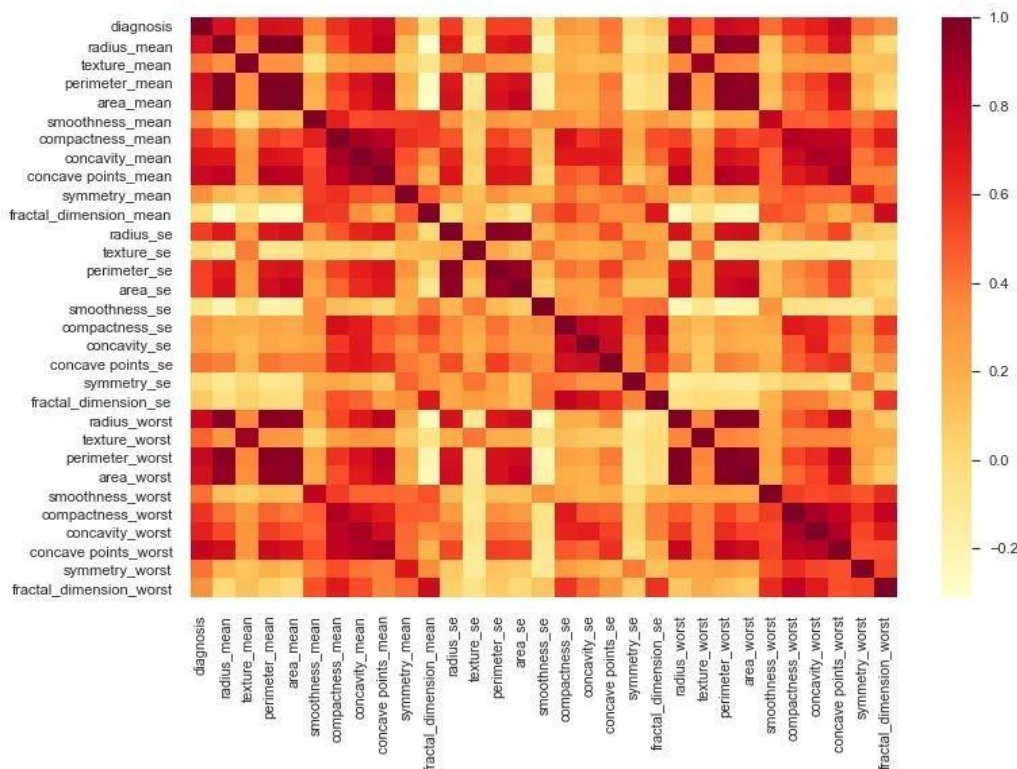**Fig.2** Sample data set from Wisconsin Medical cell

Fig.3 Correlation graph

Fig.3 Derives the Correlation graph of Breast Cancer Detection using Logistic regression, the correlation graph offers a succinct summary of the associations between predictor factors and breast cancer presence or absence. The strength and direction of the correlations between each predictor and the target variable may be determined by looking at the graph. This knowledge aids in identifying the predictors that significantly relate to breast cancer. The logistic regression model's accuracy and interpretability can be affected by multicollinearity, which can be detected using the correlation graph. Logistic regression may be used to analyze the association between predictor factors and the probability of developing breast cancer. Despite the fact that logistic regression does not by itself immediately result in a correlation graph, it can offer information about the strength and direction of the relationships between predictor variables and the likelihood of developing breast cancer.

A dataset including patient data, such as age, tumor

size, lymph node involvement, hormone receptor status, and other pertinent characteristics, is often the first step in utilizing logistic regression to analyze the association between predictor variables and breast cancer diagnosis. The direction and quantity of the predictor factors' influence on the likelihood of developing breast cancer are shown by the coefficients of the predictor variables estimated via logistic regression.

The predictor factors would be the independent variables in the logistic regression model, and the binary outcome variable would indicate whether or not breast cancer exists. The connections between the predictor variables and breast cancer may then be evaluated using the logistic regression model that is produced.

Here is an illustration of how to evaluate logistic regression results for the detection of breast cancer:

Variable	Coefficient	Odds Ratio	p-value
Age	0.5	1.65	<0.001
Tumor Size	1.2	3.32	<0.001
Lymph Node Involvement	0.8	2.23	0,0002
Hormone Receptor Status	-0.3	0.74	0.086

In this illustration, a coefficient, odds ratio, and p-value are linked to each predictor variable (age, tumor size, lymph node involvement, and hormone receptor status). For a one-unit change in the associated predictor variable, the coefficient shows the change in the log-odds of breast cancer. The odds ratio shows how changing the predictor variable by one unit multiplies the likelihood of developing breast cancer.

The direction and intensity of the relationships between the predictor factors and breast cancer may be ascertained by looking at the coefficients and odds ratios. For instance, greater tumor sizes may be linked to a higher risk of breast cancer, according to a positive coefficient and odds ratio for tumor size. On the other hand, a negative coefficient and odds ratio for the presence of hormone receptors imply that some receptor types may be linked to a lower risk of breast cancer.

The statistical significance of the computed coefficient is indicated by the p-value attached to each predictor variable. If the coefficient is substantially different from zero and the p-value is less than a preset threshold (e.g., 0.05), there is

likely a meaningful link between the predictor variable and breast cancer.

You can determine which predictor factors in breast cancer detection have the most influence by using logistic regression analysis. With the use of these insights, medical professionals may better stage breast cancer, assess prognosis, and priorities particular elements throughout the diagnostic procedure.

Logistic regression, it's vital to note, presupposes a linear connection between the predictor variables and the log-odds of breast cancer. In order to effectively represent increasingly complicated links, nonlinear relationships or interactions between variables may necessitate the use of additional modelling approaches, such as polynomial terms or interaction terms. Heat maps and scatter plots are useful visualization techniques for examining the correlation between continuous data. Correlation between two variables simply indicates that if one variable's value changes, the other one tends to change in a specific way Accuracy of testing data is the major accuracy that we are gaining in this methodology and the process

of finding the result and to support the data set result

Fig.4 Defines the precision Curve of Breast Cancer detection using Logistic regression.
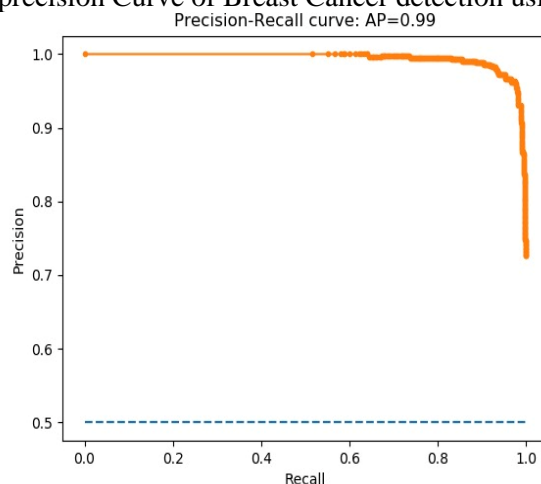

Fig.4 Precision Curve

Precision Curve of the testing data where the curve maintained at 1.0 up to 1.0 recall. The testing data's precision curve implies that the model achieves perfect precision for every positive event it detects since it maintains a precision of 1.0 up to a recall of 1.0. This shows that the model is always right when it predicts a positive case. This is the ideal situation, when the model reliably and accurately detects positive instances without producing any false positives. Many classifications jobs demand for a precision-recall curve with this level of performance, which is ideal.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	106	2
Predicted Negative (0)	0	162

Fig.5 Confusion Matrix

Confusion matrix for the Breast Cancer Detection using Logistic Regression. A confusion matrix is a useful tool for assessing how well the prediction model is working. It offers a tabular comparison of the model's forecasts and the actual classes.

V. CONCLUSION
In our research, we used machine learning models to perform a series of analyses in order to better classify breast cancer for the provided data set. As previously stated, the training dataset generated by the logistic regression method showed good results. Our model's accuracy was 92.98%. Increased accuracy and a larger data set will be possible through future work. This cancer stage prediction aids in early therapy and patient

recovery.

VI. REFERENCES
1. L. Liu, "Research on Logistic Regression Algorithm of Breast Cancer Diagnose Data by Machine Learning," 2018 International Conference on Robots & Intelligent System (ICRIS), 2018, pp. 157-160, doi: 10.1109/ICRIS.2018.00049.
2. S. Nanglia, Muneer Ahmad, Fawad Ali Khan, N.Z. Jhanjhi, An enhanced Predictive heterogeneous ensemble model for breast cancer prediction, Biomedical Signal Processing and Control, Volume 72, Part A, 2022, 103279, doi: 10.1016279.
3. ZibaKhandezamin, Marjan Naderan, Mohammad Javad Rashti, "Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier", Journal of Biomedical Informatics, Volume 111,2020,103591.
4. R. MurtiRawat, S. Panchal, V. K. Singh and Y. Panchal, "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 534-540,
doi: 10.119/ICESC48915.2020.9155783.
5. V. V. de Melo, "Breast cancer detection with logistic regression improved by features constructed by Kaizen programming in a hybrid approach," 2016 IEEE Congress on evolutionary Computation (CEC), 2016, pp. 16-23, doi:10.1109/CEC.2016.7743773.
6. A. F. Seddik and D. M. Shawky, "Logistic regression model for breast cancer automatic

diagnosis," 2015 SAI Intelligent Systems Conference (IntelliSys), 2015, pp. 150-154, doi: 10.1109/IntelliSys.2015.7361138.

7. S. Kaya and M. Yağanoğlu, "An Example of Performance Comparison of Supervised Machine Learning Algorithms Before and After PCA and LDA Application: Breast Cancer Detection," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-6,
doi: 10.1109/ASYU50717.2020.9259883.

8. A. Kumar, R. Patra, and A. Ghosh, "Model Selection for Predicting Breast Cancer using Supervised Machine Learning Algorithms," 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), 2020, pp. 320-324,
DOI: 10.1109/ICCE50343.2020.9290578.

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

10. O. Graja, M. Azam and N. Bouguila, "Breast Cancer Diagnosis using Quality Control Charts and Logistic Regression," 2018 9th International Symposium on Signal, Image, Video, and Communications (ISIVC), 2018, pp. 215-220, DOI:10.1109/ISIVC.2018.8709214.

11. P. P. Sengar, M. J. Gaikwad and A. S. Nagdive, "Comparative Study of Machine Learning Algorithms for Breast Cancer Prediction," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 796-801,
DOI: 10.1109/ICSSIT48917.2020.9214267.

12. Ajit, N &Gedam, &Deshmane, Kajol & Jadhav, Nishigandha&Adhav, Ritul&Ghodake, Akanksha &Ospanova, A.. (2022). Survey on Breast Cancer Detection using Logistic Regression Algorithm. International Journal of Special Education. 11. 3891-3895. 10.15680/IJIRSET.2022.1104129