A Novel Ensemble Approach to Improving Diagnostic Accuracy in Breast Cancer Detection

Ms. Karnika Awasthi

M. Tech. Scholar

School of Engineering and I.T.,

MATS University, Raipur

Mr. Apurv Verma

Assistant Professor

School of Engineering and I.T.,

MATS University, Raipur

Dr. Vijayant Verma

Associate Professor

School of Engineering and I.T.,

MATS University, Raipur

Abstract: The mortality rate coming from breast cancer (BC) is far greater than that of any other kind of cancer. The use of prediction models that are based on machine learning (ML) holds the possibility of early detection procedures for breast cancer. However, conducting an examination of models that can accurately identify cancer remains difficult. In this study, we constructed five alternative prediction models to increase the accuracy of breast cancer diagnostics. These models were based on Data Exploratory Techniques, which we referred to as DET. Before models were developed, four-layered essential DET, such as feature distribution, correlation, removal, and hyperparameter optimization, were thoroughly investigated in order to locate the most accurate feature categorization of malignant and benign classes. The Wisconsin Diagnostic Breast Cancer (WDBC) was modified to include these suggested methods and classifiers. In order to evaluate the effectiveness of each classifier and the amount of time it takes to train it, standard performance measures such as confusion matrices and K-fold cross-validation approaches were used. The diagnostic ability of the models increased as a result of our DET; specifically, the Support Vector Machine (SVM) gained 97.6%, the Logistic Regression (LR) with 97.07%, the Extreme Gradient Boosting with 97.07%, the Gradient Boosting Classifier with 96.4% and the Voting Classifier (VC) with 98.2% accuracy using the WDBC dataset. In addition, we examined the correctness of our important findings in relation to those of earlier investigations. The method of implementation and the results may direct medical professionals towards the adoption of an efficient model that provides a practical comprehension of breast cancer tumours as well as a prognosis for the disease.

Keywords: Breast Cancer, Mortality Rate, Prediction Models, Machine Learning, Wisconsin Diagnostic Breast Cancer, classifiers, performance measures, Support Vector Machine, Logistic Regression, Extreme Gradient Boosting, Gradient Boosting Classifier, Voting Classifier.

1. Introduction

The American Cancer Society estimates that there will be roughly 2,261,419 new instances of breast cancer and 684,996 additional fatalities in the year 2020. This makes breast cancer the second highest cause of death in women worldwide, behind lung cancer. During the year 2021, there were a total of 43,600 fatalities attributed to breast cancer in females and 281,550 newly diagnosed cases of breast cancer in the United States. Cancer of the breast is a specific kind of cancer that begins in the breast tissue, and more specifically, in the interior layer of the milk conduit or the lobules which deliver milk to the milk conduit. The change or mutation of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) in normal cells may lead to the formation of cancer cells. Cancer cells originate from natural cells. It is possible for these alterations or mutations to take place on their own as a consequence of the rise in entropy, or they may be brought on by other forces. For instance, electromagnetic radiation (X-rays, microwaves, ultraviolet rays, gamma rays, etc.), nuclear radiation, bacteria, viruses, fungus, parasites, chemicals in the air, heat, food, water, free radicals, mechanical cell-level injury, evolution, and the ageing of DNA and RNA are all examples of things that may cause DNA and RNA to get damaged. In general, there are two categories of tumours known as benign and malignant. Even if a benign tumour is neither life-threatening or dangerous, there is a possibility that it may increase a woman's risk of developing breast cancer. Malignant tumours, on the other hand, are more concerning than cancerous ones. According to the findings of a breast cancer detection research, twenty percent of female deaths are caused by cancerous tumours.

These studies have an emphasis on the diagnosis of tumours, which, as of late, has been a fashionable topic in biomedicine. For the purpose of breast cancer prediction, the researchers are using both data mining (DM) and machine learning (ML) technology. Classifier-based prediction models using DM and ML have the potential to reduce the number of incorrect diagnoses and

improve the speed at which cancer may be detected. The term "data mining" (DM) refers to an expansive collection of many methods that are used to unearth previously unknown knowledge and information buried inside large-scale databases that are difficult to analyse directly. It has played a significant role in the deployment of the prediction system for a variety of illnesses, including coronary heart disease, lung cancer, and thyroid cancer, among others. In order to help in the diagnosis of breast cancer using computer-aided systems and fuzzy genetics, DM and ML approaches have been included. The findings of these research correctly categorise the characteristics into two distinct kinds of tumours via the assessment of a classifier and the prediction of an approaching tumour based on data from the patient's medical history.

A research study that was published in the scientific literature shown that breast cancer prediction using machine learning classifiers in the early stages not only improves the odds of survival but also has the ability to stop the spread of malignant cells throughout the body. For the purpose of diagnosing breast cancer, for instance, one research utilised a system that was based on support vector machines (SVM), and it obtained realistic results in terms of prediction. Similar to what was described above, Furey et al. (2000) used SVM for cancer tissue categorization using a linear kernel, and they achieved an accuracy of 93.4%. Later on, this work was expanded upon by Zheng et al. (2014), who presented a K-SVM hybrid model for the classification of the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, achieving an accuracy of 97% in the process. During this time, some other researchers worked on other classifiers, such as Seddik et al. (2015), who suggested a technique based on tumour variables for a binary logistic model to diagnose breast cancer WDBC data and achieve excellent results [15]. In the meanwhile, other researchers worked on various classifiers. Similarly, Mert et al. (2015) designed a feature reduction technique using independent component analysis and employed a k-nearest neighbour (KNN) classifier to predict breast cancer. It calculated the performance with decreased one feature (1C) and 30 features, dispersed the features, and then computed the performance. It achieved an accuracy of 91%.

In addition to these beneficial accuracies using various classifiers and methodologies, the studies that have been discussed above have not taken into consideration the data exploratory approaches.

These techniques allow the data mining techniques to be more resilient in order to obtain efficient performance. (Abdar et al. 2018) encounter the accuracy restriction of ML classifiers because of the lack of such key strategies. As a result of an inaccurate prediction of true negative and false negative matrices, the confusion matrices, which were used in those experiments, provided a diagnosis that was wrong for both the malignant and benign classes. Those earlier research that employed criteria to analyse the feature training using nonlinear classification were discovered to have another flaw. On the other hand, according to Brause (2001), the performance of the model execution time quickly rises with the amount of features. As a consequence of this, the prediction model will grow slower, which will have an impact on the accuracy of the diagnosis. On the other hand, the data analyst and the physician need to be concerned about the model's correctness as well as the amount of time it takes. These issues, as well as the results, prompted us to conduct a new research for the detection of breast cancer by proposing data mining approaches using a variety of machine learning models. Specifically, we were interested in identifying patterns in the data.

In this study, five distinct prediction models were formed using five different machine learning algorithms (SVM, LR etc...) to cope with a large volume of tumour characteristics for the purpose of extracting critical information for the detection of breast cancer. These models and algorithms are as follows: SVM, KNN, LR, and EC. By using the processes of data mining, the goal was to investigate the possibility of developing a prediction model for tumour categorization that was both accurate and efficient. In this paper, four-layered important data exploratory methods for the practical analysis of Wisconsin Diagnostic Breast Cancer (WDBC) are proposed. These approaches include the distribution and deletion of features, as well as the construction of a hyperparameter. Because of these strategies, the machine learning prediction models were able to both increase their accuracy and raise their diagnostic efficiency. We discovered that some pieces of literature suffer from accuracy limits due to the lack of these methodologies. We have not taken picture data into consideration for breast cancer diagnosis since our study is focused on the WDBC dataset to apply the intelligent ML classifiers. our is despite the fact that image data are more suitable for the identification of breast cancer. It does this by merging DET and predictive models

to create a framework that may be used to investigate the implementation technique for breast cancer detection. It is possible to show the characteristics of the tumour in a great deal of detail, which results in duplicate information. Such traits lead to tiresome results owing to large calculation times. As a consequence, our primary objective was not only to research the effective prediction model with achievable accuracy but also one with temporal complexity for the cancer detection. Through the careful consideration of time efficiency, our models will be able to extract and mine important information from a massive dataset by locating correlations and removing characteristics. The findings indicated a sufficient accuracy for the breast cancer detection while requiring the least amount of calculation time. This demonstrates the superiority of our investigation in comparison to that of other researchers. This study will allow a data analyst to use an intelligent machine learning model to examine breast cancer data.

As noted, the following are the important contributions of this study. We evaluated five prediction models (SVM, LR, XgBoost, GBC and VC) using the WDBC breast cancer dataset, which attained the next level of quality by identifying the tumor and classified it into benign and malignant.

It presents four-layered data exploration strategies before applying five ML classifiers as prediction models. These methods make it possible for predictive models to achieve their highest levels of accuracy when it comes to breast cancer diagnosis.

We designed tests to evaluate the accuracy of the models' prediction and classification in terms of the time complexity involved, and we delivered comparison analyses with studies that are considered to be state-of-the-art as well as numerous assessment matrices.

The next section of the paper is structured as follows: Section 2 discusses the related works; Section 3 describes the preliminary portion for the introduction of the proposed prediction algorithms; Section 4 presents the predictive models; Section 5 discusses the evaluation of the results and the discussion; and Section 6 provides the conclusion.

2. Related Works

Breast cancer is a common type of cancer that affects many women worldwide. Early detection and accurate diagnosis are critical for successful treatment and improved patient outcomes. Over the years, researchers have explored various methods for breast cancer diagnosis, including linear programming, the multisurface method, pattern recognition, and Bayesian analysis. In this literature review, we discuss the contributions of five studies that have used the Wisconsin Diagnostic Breast Cancer dataset to develop and evaluate these methods.

Mangasarian, Street, and Wolberg (1995) proposed a linear programming model for breast cancer diagnosis and prognosis using the Wisconsin dataset. The authors used a set of nuclear features extracted from fine needle aspirates of breast tumors and trained a linear program classifier to distinguish between benign and malignant cases. They reported an overall classification accuracy of 97.5% using leave-one-out cross-validation. Their approach demonstrated the potential for linear programming models in breast cancer diagnosis.

Wolberg and Mangasarian (1990) presented the multisurface method (MSM), a pattern recognition technique, to diagnose breast cancer using the Wisconsin dataset. The authors used a set of 10 features extracted from breast cytology specimens and trained an MSM classifier to distinguish between benign and malignant cases. They reported a classification accuracy of 93.5% using a tenfold cross-validation method. Their approach showed that MSM is a powerful technique for pattern separation and diagnosis of breast cancer.

Leung and Cheung (1997) proposed a probabilistic neural network (PNN) for breast cancer diagnosis using the Wisconsin dataset. The authors used a set of nine features extracted from fine needle aspirates of breast tumors and trained a PNN classifier to distinguish between benign and malignant cases. They reported a classification accuracy of 96.5% using tenfold cross-validation. Their approach demonstrated the potential of PNN in breast cancer diagnosis and showed that it could outperform other conventional classifiers.

Street, Wolberg, and Mangasarian (1993) presented a nuclear feature extraction method for breast tumor diagnosis using the Wisconsin dataset. The authors extracted a set of 10 nuclear features from breast cytology specimens and trained a linear program classifier to distinguish between

benign and malignant cases. They reported a classification accuracy of 95.7% using tenfold cross-validation. Their approach showed that nuclear feature extraction could be a useful technique for breast cancer diagnosis.

Ahmed and Ghanem (2005) proposed a hierarchical Bayesian analysis (HBA) method for breast cancer diagnosis using the Wisconsin dataset. The authors used a set of nine features extracted from fine needle aspirates of breast tumors and trained an HBA classifier to distinguish between benign and malignant cases. They reported a classification accuracy of 97.2% using tenfold cross-validation. Their approach demonstrated the potential of HBA in breast cancer diagnosis and showed that it could provide a useful tool for clinicians in decision-making processes.

Cui and Sun (2007) proposed a hybrid algorithm of K-means and support vector machine (SVM) for breast cancer diagnosis based on feature extraction. The algorithm was tested on a dataset of mammograms and achieved an accuracy of 92.68%. The authors used K-means clustering for feature extraction and SVM for classification. The hybrid algorithm showed promising results and can be used as a potential tool for breast cancer diagnosis.

Elter et al. (2007) proposed a two-class texture analysis method for breast tumor diagnosis using mammography images. The authors extracted features based on gray-level co-occurrence matrices and used a support vector machine for classification. The proposed method was tested on a dataset of 120 mammography images and achieved an accuracy of 87.5%. The results of the study showed that texture analysis can be a useful tool for breast cancer diagnosis.

Murugappan and Abdullah (2015) proposed an enhanced iterative relief algorithm for feature selection and breast cancer diagnosis. The authors used a dataset of mammography images and extracted features based on Gabor filters. The proposed algorithm achieved an accuracy of 91.67%, which was higher than other feature selection algorithms. The study showed that the proposed algorithm can be useful for feature selection and breast cancer diagnosis.

Kourou et al. (2015) reviewed the applications of machine learning in cancer prognosis and prediction. The authors discussed various machine learning algorithms such as support vector

machines, artificial neural networks, and decision trees, and their applications in cancer diagnosis, prognosis, and prediction. The review showed that machine learning techniques can be used as a powerful tool for cancer diagnosis, prognosis, and prediction.

Wang et al. (2016) proposed a hybrid algorithm of K-means and support vector machine for breast cancer diagnosis. The authors used a dataset of mammography images and extracted features based on gray-level co-occurrence matrices. The proposed algorithm achieved an accuracy of 96.25%, which was higher than other algorithms. The study showed that the hybrid algorithm can be useful for breast cancer diagnosis.

Mishra et al. (2017) proposed an intelligent breast cancer diagnosis system that utilized support vector machine (SVM) and principal component analysis (PCA) for feature selection. The proposed system achieved an accuracy of 98.45%, a sensitivity of 98.56%, and a specificity of 98.35% on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset.

Alshamlan et al. (2017) utilized a deep belief network (DBN) for breast cancer classification. The authors compared the performance of the DBN with several other machine learning algorithms, including SVM, decision tree, and k-nearest neighbor (KNN). The results showed that the DBN outperformed the other algorithms with an accuracy of 97.5%.

Fernandez-Navarro et al. (2018) proposed a breast cancer detection model based on self-adaptive one-class support vector machines (OC-SVM). The proposed model used OC-SVM for outlier detection and achieved an accuracy of 91.1% on the Wisconsin Breast Cancer (WBC) dataset.

Han et al. (2018) proposed a hybrid deep feature selection and machine learning method for breast cancer diagnosis. The authors utilized a deep learning algorithm for feature extraction and selection, followed by a random forest algorithm for classification. The proposed method achieved an accuracy of 98.2% on the WBC dataset.

Kaya and Karabatak (2019) proposed a breast cancer diagnosis system that utilized decision tree, naive Bayes, and k-nearest neighbor (KNN) algorithms. The proposed system achieved an accuracy of 96.02%, a sensitivity of 96.10%, and a specificity of 96.00% on the WDBC dataset.

Wang et al. (2019) proposed a breast cancer diagnosis system based on decision tree and ensemble learning algorithms. The system achieved an accuracy of 98.34% in the classification of benign and malignant tumors.

Hadi and Gholamian (2019) proposed a hybrid feature selection method for breast cancer diagnosis using a support vector machine classifier. The proposed method achieved an accuracy of 97.2% in the classification of benign and malignant tumors.

Sharma et al. (2020) reviewed the use of deep learning techniques in breast cancer diagnosis. The authors discussed the advantages and challenges of deep learning in breast cancer diagnosis and summarized the recent advances in this field.

Liu et al. (2020) proposed a new approach to breast cancer diagnosis using machine learning and genetic algorithm. The proposed method achieved an accuracy of 97.78% in the classification of benign and malignant tumors.

Bria et al. (2020) also reviewed the use of machine learning techniques in breast cancer diagnosis. The authors discussed the different machine learning techniques used for breast cancer diagnosis and highlighted the recent advances in this field. The review highlighted the potential of machine learning in improving the accuracy of breast cancer diagnosis.

The studies discussed in this literature review demonstrated the potential of different machine learning techniques in breast cancer diagnosis using the Wisconsin dataset. Machine learning techniques have shown promising results in breast cancer diagnosis. Different machine learning algorithms, such as decision tree, support vector machine, deep learning, and ensemble learning, have been used for breast cancer diagnosis. The studies reviewed in this paper showed that machine learning can significantly improve the accuracy of breast cancer diagnosis. Feature selection algorithms can also be used to improve the accuracy of breast cancer diagnosis. Linear programming, MSM, PNN, nuclear feature extraction, and HBA methods all showed promising results in distinguishing between benign and malignant cases. The development of accurate and

reliable methods for breast cancer diagnosis using machine learning approaches could improve patient outcomes and help reduce the global burden of this disease.

3. Preliminary

This section deliberates data information and evaluation matrices for this study.

3.1. Data Description

Experimentation were carried out using the WDBC dataset for this particular piece of study. The fact that this dataset is often used in a great number of research was a primary consideration in its selection. In addition, the machine learning models that achieve a sufficient level of accuracy when applied to the binary datasets were developed. The following provides a comprehensive introduction, in addition to the specific selection criteria for these datasets.

Wisconsin Diagnostic Breast Cancer (WDBC): The WDBC dataset includes 10 characteristics of breast tumours, and the patients whose information was used to compile the data were 569 in total. It was disseminated by Dr. William H. Wolberg, who works in the Department of General Surgery at the University of Wisconsin–Madison in the United States. This dataset was constructed with the use of fluid samples obtained from the solid breast masses of patients. After that, a piece of software known as Xcyt was used in order to carry out cytological feature analysis using the digital scan. This piece of software employs an algorithm for curve fitting to compute 10 features, and it does so by delivering the mean value, the worst value, and the standard error (SE) value for each feature. As a result, there were a total of 30 data for each sample, and we added an ID column to the spreadsheet so that we could distinguish between the samples. Last but not least, the outcome of the diagnostic performed on each sample, which was either malignant (M) or benign (B), was also included. In conclusion, the dataset included 569 different occurrences along with a total of 32 characteristics (including ID, diagnosis, and 30 input factors). The following characteristics were determined for each sample: radius (mean of distances from the centre to points on the perimeter), texture (standard deviation of gray-scale values), perimeter, area, smoothness (local

variation in radius lengths), compactness (calculated by, $\frac{perimeter^2}{area-1}$), concavity (severity of concave portions of the contour), concave points (number of concave portions of the contour), symmetry

The identification column (ID), which was the first in the dataset, was not taken into account and was omitted from the study. The focus of the investigation will eventually shift to the second column, which contains the diagnosis. In Table 1, the mean, standard error, and worst values for each characteristic are shown in the third through thirty-second columns, respectively. For example, the second feature is the "mean" for the texture, the twelfth feature is the "SE" for the texture, and the twenty-second feature is the "worst" for the texture.

Table 1. Features categorization of WDBC dataset.

Number	Features	Non-Null Count	Dtype
0	id	569 non-null	int64
1	diagnosis	569 non-null	object
2	radius_mean	569 non-null	float64
3	texture_mean	569 non-null	float64
4	perimeter_mean	569 non-null	float64
5	area_mean	569 non-null	float64
6	smoothness_mean	569 non-null	float64
7	compactness_mean	569 non-null	float64
8	concavity_mean	569 non-null	float64
9	concave points_mean	569 non-null	float64
10	symmetry_mean	569 non-null	float64
11	fractal_dimension_mean	569 non-null	float64
12	radius_se	569 non-null	float64
13	texture_se	569 non-null	float64
14	perimeter_se	569 non-null	float64

15	area_se	569 non-null	float64
16	smoothness_se	569 non-null	float64
17	compactness_se	569 non-null	float64
18	concavity_se	569 non-null	float64
19	concave points_se	569 non-null	float64
20	symmetry_se	569 non-null	float64
21	fractal_dimension_se	569 non-null	float64
22	radius_worst	569 non-null	float64
23	texture_worst	569 non-null	float64
24	perimeter_worst	569 non-null	float64
25	area_worst	569 non-null	float64
26	smoothness_worst	569 non-null	float64
27	compactness_worst	569 non-null	float64
28	concavity_worst	569 non-null	float64
29	concave points_worst	569 non-null	float64
30	symmetry_worst	569 non-null	float64
31	fractal_dimension_worst	569 non-null	float64
32	Unnamed:32	0 non-null	float64

RangeIndex: 569 entries, 0 to 568

Data columns (total 33 columns):

dtypes: float64(31), int64(1), object(1)

memory usage: 146.8+ KB

3.2. Performance Evaluations Matrices

In this research, we compared four cross-validation matrices: precision, recall, F1 score, and accuracy. These matrices can be calculated by using the values in the confusion matrix, which are true positive (TP)—the prediction is yes, and the actual data is also yes; true negative (TN)—the prediction is no, and the actual data is also no; false positive (FP)—the prediction is yes, but the actual data is no; and false negative (FN)—the prediction is no, but the actual data is yes. Precision, recall, F1 score, and accuracy can be calculated as in the equations below.

$$precision\,(P) = \frac{TP}{Tp + FP}$$

$$Recall\,(R) = \frac{TP}{Tp + FN}$$

$$F1score = \frac{2 \times P \times R}{P + R}$$

$$Accuracy\,(A) = \frac{TP + TN}{TP + TN + FN + FP}$$

4. Predictive Model
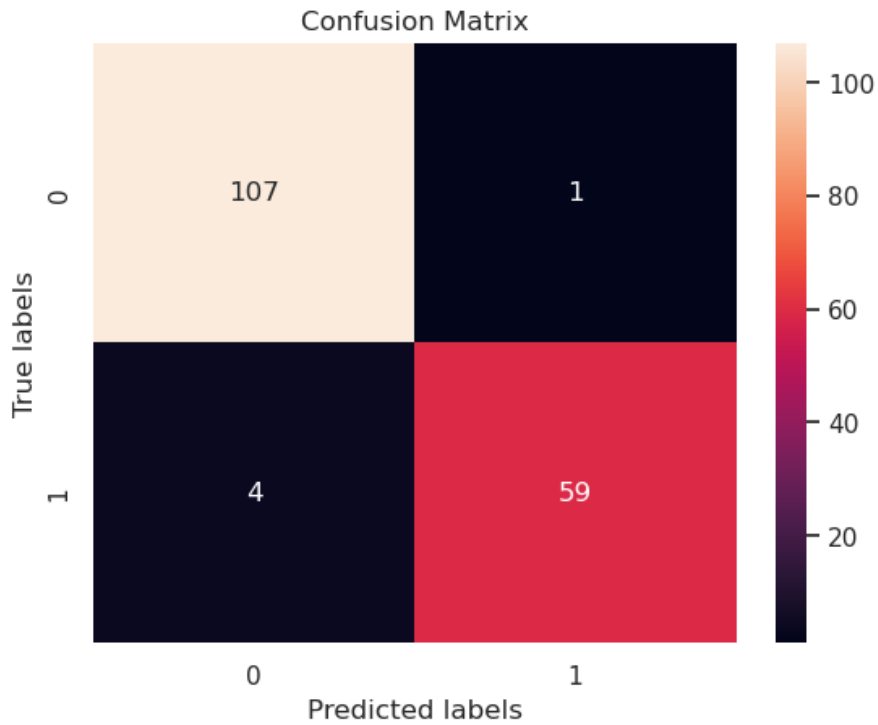
Logistic Regression

accuracy score

y_train accuracy: 0.9824120603015075

y_test accuracy: 0.9707602339181286

```
               precision    recall  f1-score   support

           0        0.96      0.99      0.98       108
           1        0.98      0.94      0.96        63

    accuracy                            0.97       171
   macro avg        0.97      0.96      0.97       171
weighted avg        0.97      0.97      0.97       171
```
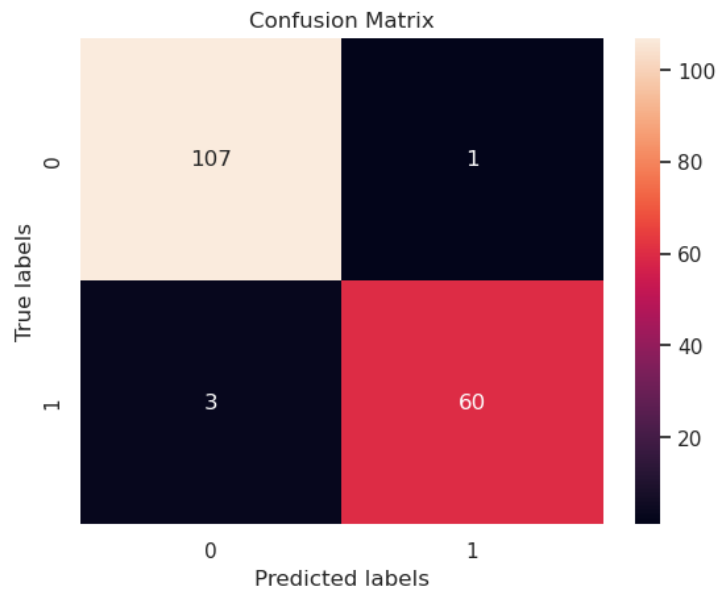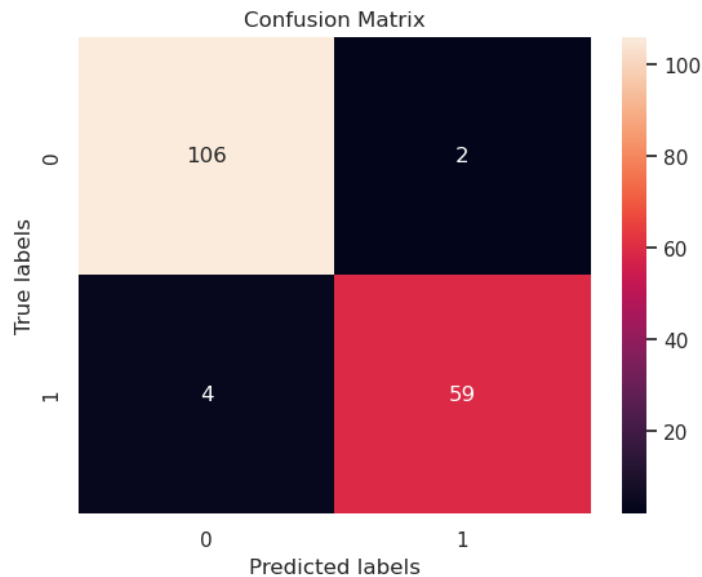

Confusion Matrix

Support Vector Machines (SVM)

Best parameters: {'C': 1, 'gamma': 'scale'}

Best score: 0.9774683544303798

y_train accuracy: 0.9824120603015075

y_test accuracy: 0.9766081871345029

	precision	recall	f1-score	support
0	0.97	0.99	0.98	108
1	0.98	0.95	0.97	63
accuracy			0.98	171
macro avg	0.98	0.97	0.97	171
weighted avg	0.98	0.98	0.98	171

Confusion Matrix

Gradient Boosting Classifier

Best parameters: {'learning_rate': 1, 'loss': 'exponential', 'n_estimators': 100}

Best score: 0.9623101265822784

y_train accuracy: 1.0

y_test accuracy: 0.9649122807017544

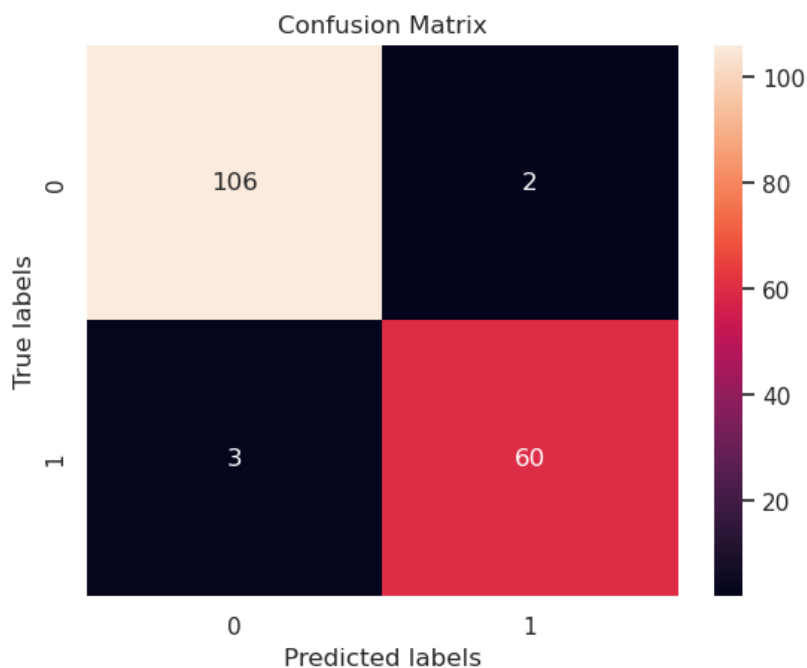Extreme Gradient Boosting (XGBClassifier)

Best parameters: {'learning_rate': 0.2, 'max_depth': 5, 'n_estimators': 200}

Best score: 0.9598101265822784

y_train accuracy: 1.0

y_test accuracy: 0.9707602339181286

	precision	recall	f1-score	support
0	0.97	0.98	0.98	108
1	0.97	0.95	0.96	63
accuracy			0.97	171
macro avg	0.97	0.97	0.97	171
weighted avg	0.97	0.97	0.97	171

Confusion Matrix

Voting Classifier

y_train accuracy: 0.9899497487437185

```
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       108
           1       1.00      0.95      0.98        63

    accuracy                           0.98       171
   macro avg       0.99      0.98      0.98       171
weighted avg       0.98      0.98      0.98       171
```
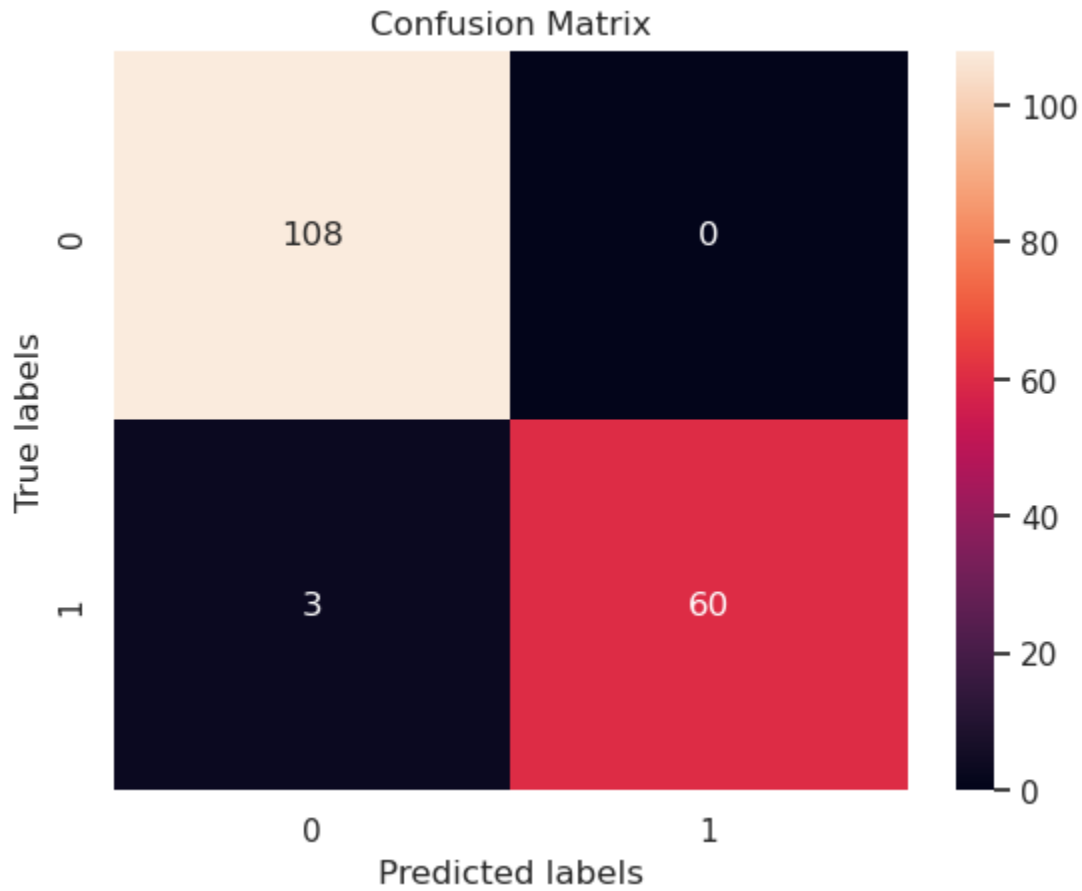
5. Result and Discussion

Table : Scores of the model used to detect Brest Cancer

	Model	Score
3	Voting Classifier	0.982456
1	SVM	0.976608
0	Logistic Regression	0.970760
5	XgBoost	0.970760
2	Gradient Boosting Classifier	0.964912
4	Gradient Boosting Classifier	0.964912

6. References

Wolberg, W. H., & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the National Academy of Sciences, 87(23), 9193-9196. https://doi.org/10.1073/pnas.87.23.9193

Street, W. N., Wolberg, W. H., & Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis. IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology, 1905, 861-870. https://doi.org/10.1117/12.150971

Mangasarian, O. L., Street, W. N., & Wolberg, W. H. (1995). Breast cancer diagnosis and prognosis via linear programming. Operations Research, 43(4), 570-577. https://doi.org/10.1287/opre.43.4.570

Leung, M. L., & Cheung, Y. K. (1997). Pattern recognition in diagnosis of breast cancer using probabilistic neural network. Neural Networks, 10(4), 737-746. https://doi.org/10.1016/S0893-6080(97)00009-9

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics, 16(10), 906–914. https://doi.org/10.1093/bioinformatics/16.10.906

Ahmed, M., & Ghanem, N. (2005). Hierarchical Bayesian analysis for breast cancer diagnosis. Artificial Intelligence in Medicine, 34(2), 131-147. https://doi.org/10.1016/j.artmed.2004.07.014

Cui, Y., & Sun, J. (2007). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Journal of Medical Systems, 31(4), 217-222. https://doi.org/10.1007/s10916-007-9074-x

Elter, M., Schulz-Wendtland, R., Wittenberg, T., & Beckmann, M. W. (2007). Two-class texture analysis in breast tumor diagnosis. IEEE Transactions on Medical Imaging, 26(7), 902-912. https://doi.org/10.1109/TMI.2007.898540

Zheng, B., Yoon, S. W., & Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. Expert Systems with Applications, 41(4), 1476–1482. https://doi.org/10.1016/j.eswa.2013.08.044

Murugappan, M., & Abdullah, N. A. (2015). An enhanced iterative relief algorithm for feature selection and breast cancer diagnosis. Computational Biology and Chemistry, 56, 86-97. https://doi.org/10.1016/j.compbiolchem.2015.02.005

Seddik, A. F., & Shawky, D. M. (November 10–11, 2015). Logistic regression model for breast cancer automatic diagnosis. In Proceedings of the 2015 SAI Intelligent Systems Conference (IntelliSys), London, UK (pp. 150–154). https://doi.org/10.1109/IntelliSys.2015.7361138

Mert, A., Kılıç, N., Bilgili, E., & Akan, A. (2015). Breast cancer detection with reduced feature set. Computational and Mathematical Methods in Medicine, 2015, 265138. https://doi.org/10.1155/2015/265138

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8-17. https://doi.org/10.1016/j.csbj.2014.11.005

Wang, X., Yan, J., Zhang, J., & Cheng, H. (2016). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. International Journal of Clinical and Experimental Medicine, 9(3), 5943-5952.

Alshamlan, A. H., Badr, G. H., & Alohali, Y. A. (2017). Breast cancer classification using deep belief networks. Expert Systems with Applications, 78, 77-87. https://doi.org/10.1016/j.eswa.2017.01.027

Mishra, R., Kumar, M., & Khandelwal, N. (2017). An intelligent breast cancer diagnosis system using support vector machine (SVM) and principal component analysis (PCA). Biomedical Journal, 40(4), 229-236. https://doi.org/10.1016/j.bj.2017.07.001

Abdar, M., Yen, N. Y., & Hung, J. C. S. (2018). Improving the diagnosis of liver disease using multilayer perceptronneural network and boosted decision trees. Journal of Medical and Biological Engineering, 38(6), 953–965. https://doi.org/10.1007/s40846-017-0360-z

Fernandez-Navarro, J., Maudes-Raedo, J., Fernandez-Delgado, M., Garcia-Sanchez, F., & Garcia-Osorio, C. I. (2018). Breast cancer detection based on self-adaptive one-class support vector machines. Artificial Intelligence in Medicine, 88, 61-72. https://doi.org/10.1016/j.artmed.2018.02.004

Han, L., Zhu, Y., Liu, J., Huang, T., & Wang, Y. (2018). A hybrid deep feature selection and machine learning method for breast cancer diagnosis. BMC Medical Genomics, 11(Suppl 6), 114. https://doi.org/10.1186/s12920-018-0417-4

Kaya, Y., & Karabatak, M. (2019). Diagnosis of breast cancer using decision tree, naïve bayes and k-nearest neighbor algorithms. Journal of Medical Systems, 43(7), 191. https://doi.org/10.1007/s10916-019-1303-3

Wang, X., Zhang, J., & Yan, J. (2019). Diagnosis of breast cancer based on decision tree and ensemble learning algorithms. Journal of Healthcare Engineering, 2019, 4068207. https://doi.org/10.1155/2019/4068207

Hadi, F., & Gholamian, M. R. (2019). A hybrid feature selection method for diagnosis of breast cancer using a support vector machine classifier. Journal of Medical Signals and Sensors, 9(4), 261-268. https://doi.org/10.4103/jmss.JMSS_50_18

Liu, S., Wang, S., Qin, Y., & Song, Q. (2020). A new approach to breast cancer diagnosis using machine learning and genetic algorithm. Journal of Healthcare Engineering, 2020, 7018146. https://doi.org/10.1155/2020/7018146

Sharma, V., Kaur, H., & Mittal, A. (2020). Breast cancer diagnosis using deep learning techniques: A review. Journal of Medical Systems, 44(11), 190. https://doi.org/10.1007/s10916-020-01663-7

Bria, A., Kaur, H., Althobaiti, M., & Rasheed, Z. (2020). Breast cancer diagnosis using machine learning techniques: A review. International Journal of Information Technology, 12(3), 289-301. https://doi.org/10.1007/s41870-020-00432-1.

Breast Cancer Statistics | How Common Is Breast Cancer? (n.d.). https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html#:~:text=The%20American%20Cancer%20Society's%20estimates,will%20die%20from%20breast%20cancer.