Overall, it appears that hyperparameter tuning successfully improved the accuracy of both AdaBoost and XGBoost, with AdaBoost seeing the greatest improvement. It is possible that further improvements could be achieved through additional hyperparameter tuning or by trying different algorithms.
A confusion matrix makes it simple to summarize the performance of a classification method. Calculating a confusion matrix will help us better understand.

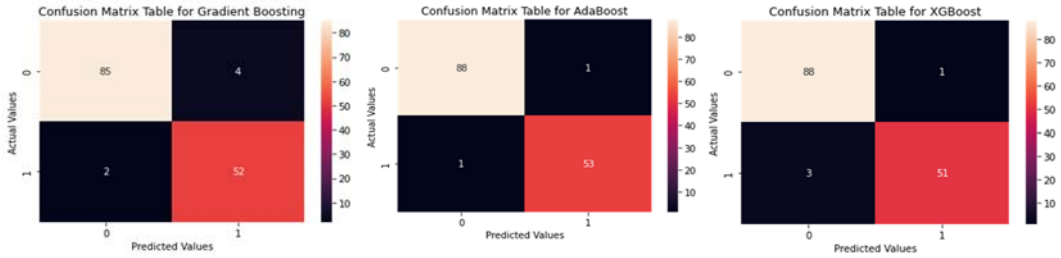


Figure 03: Confusion Matrix After Tuning

Area Under the Curve (AUC) calculates the two-dimensional area under the entire Receiver Operating Characteristic (ROC) curve ranging from (0,0) to (1,1).

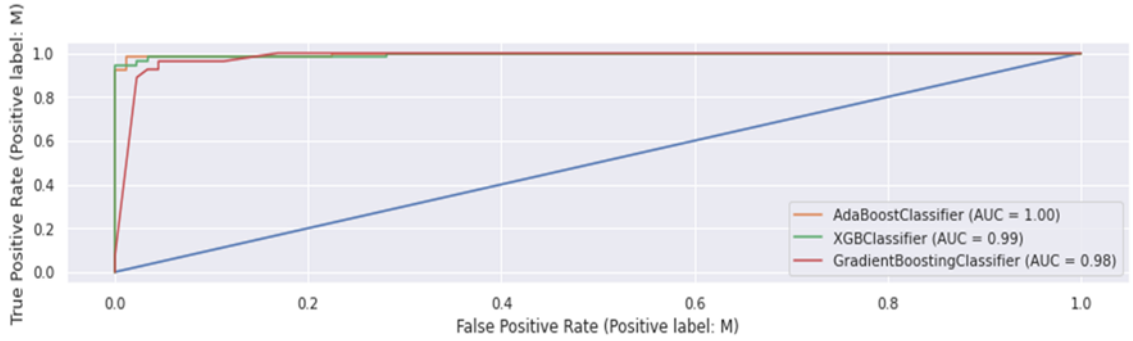


Fig 04: AUC & ROC Curve After Tuning

Our findings indicated that the suggested model AdaBoost Classifier had the highest AUC, which was 1.00. and XGBoost and Gradient Boosting had AUCs of 0.99 and 0.98.

## 4.2 Discussion

The study results indicate that AdaBoost and XGBoost performed better than Gradient Boosting in terms of accuracy. Afer hyperparameter tuning implementation, AdaBoost had the highest accuracy of 98.6%, followed by XGBoost with 97.2% and Gradient Boosting with 95.8%. In terms of other performance metrics such as precision, specificity, and recall, AdaBoost and XGBoost also performed better than Gradient Boosting. With respect to the confusion matrix, AdaBoost had the highest number of true positives and true negatives, indicating that it had the highest number of correct predictions. It also had the lowest number of false positives and false negatives, indicating that it had the lowest number of incorrect predictions. Overall, the results suggest that AdaBoost is the most effective machine-learning algorithm for this particular study. However, it is important to note that the performance of any machine learning algorithm depends on various factors, such as the quality of the data, the choice of features, and the hyperparameter tuning. Therefore, these results should be interpreted with caution, and further studies may be needed to confirm these findings.
In table 03, we made a comparison table for available prediction machine-learning models on breast cancer.

Table: 03 Comparison of publicly available prediction model

| Ref No | Period | Dataset Name | Algorithm Name | Accuracy (%) |
|---|---|---|---|---|
| [5] | 2020 | UCI BCD | RF, XGBoost | 74.73%,73.63% |
| [7] | 2021 | WBCD | CatBoost, XGBoost, DT, KNN | 97.80%,97.08%,95.60%,97% |
| [15] | 2021 | WBCD | DT, AdaBoost, | 90.20%,96.50% |
| [16] | 2022 | WBCD | GB, KNN | 95.34%,75.96% |
| [17] | 2022 | WBCD | XGB, AdaBoost | 98.24%,94.73% |
| [18] | 2020 | WBCD | AdaBoost, GB | 96.81%,97.34% |
| [20] | 2020 | WBCD | GB, XGBoost, AdaBoost | 95.96%,97.19%,95.96% |
| Our paper | 2022 | WBCD | GB<br>XGBoost<br>AdaBoost | 95.80%<br>97.20%<br>98.60% |

In this research, the performance of several boosting classifiers was evaluated on the WBCD dataset. The classifiers used were Decision Tree, AdaBoost, Gradient Boosting, XGBoost, CatBoost, and KNN. The accuracy of each classifier was measured and recorded. The results show that XGBoost and AdaBoost had the highest accuracy rates, with 98.60% and 97.20% respectively. GB also performed well, with a 95.80% accuracy rate. Compared with previous research results on the WBCD dataset, XGBoost and AdaBoost performed better than the other classifiers in this dataset. GB also performed similarly well compared with a 95.34% and 97.34% accuracy rate of other previous research. In conclusion, the results of this research suggest that AdaBoost is an effective classifier for the WBCD dataset to predict breast cancer.

## 5. Conclusions

The study aims to distinguish between malignant and benign patients, which can be highly helpful for patients and doctors in prescribing the right medications. The suggested approach can serve as an alternative to the current testing requirements and aid in the early diagnosis of breast cancer. Additionally, a side-by-side comparison is shown in this study, and the best classifier for the model that offers dependability is selected. With this research paper, we can see that among Gradient Boosting, XGBoost, and AdaBoost, the AdaBoost is the most accurate algorithm for the best accurate result for the detection of breast cancer type with an efficiency of 98.60%. In the medical industry, the diagnosing process is very time-consuming. The system proposed that a clinical assistant may be used to diagnose breast cancer using machine learning techniques.

In the future, to attain high accuracy, we plan to parametrize our categorization methods. In this study, our dataset is the shortest, but our research could have been done with a larger dataset, so we are investigating a variety of datasets and the potential applications of machine learning techniques to describe breast cancer further and want to maximize accuracy with a larger dataset while lowering error rates.

## Conflict of interest

All authors declare no conflicts of interest in this paper.

## 6. References

[1] Key, T. J., Verkasalo, P. K., & Banks, E. Epidemiology of breast cancer. The lancet oncology, 2(3), 133-140, 2001.
[2] U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2008 Incidence and Mortality Web-based Report. Atlanta (GA): Department of Health and Human Services, National Cancer Institute; 2012.
[3] Turkki, R.; Byckhov, D.; Lundin, M.; Isola, J.; Nordling, S.; Kovanen, P.E.; Verrill, C.; von Smitten, K.; et al. Breast cancer outcome prediction with tumor tissue images and machine learning. Breast Cancer Res. Treat., 177, 41–52, 2019. [CrossRef] [PubMed]
[4] S. Gokhale., "Ultrasound characterization of breast masses", The Indian journal of radiology & imaging, Vol. 19, pp. 242-249, 2009.
[5] Sajib Kabiraj,M. Raihan,Nasif Alvi,Marina Afrin,Laboni Akter,Shawmi Akhter Sohagi,Etu Podder. Breast Cancer Risk Prediction using XGBoost and Random Forest Algorithm. 10.1109/ICCCNT49239.2020.9225451, 2020.
[6] Kurian, Babymol & Jyothi, V.L. Comparative Analysis of Machine Learning Methods for Breast Cancer Classification in Genetic Sequences. Journal of Environmental and Public Health. 1-6. 10.1155/2022/7199290, 2022.
[7] H. Gupta, P. Kumar, S. Saurabh, K. Mishra, B. Appasani, A. Pati, C. Ravariu, A. Srinivasulu. Category Boosting Machine Learning Algorithm for Breast Cancer Prediction. Vol. 66, 3, pp. 201–206, Bucarest, 2021.
[8] M. Amrane, S. Oukid, I. Gagawa, and T. Ensari, “Breast cancer classification using machine learning,” in 2018 Electric Electronics, Computer Science, Biomedical Engineerings’ Meeting (EBBT), pp. 1–4,2018.
[9] Sakri SB, Rashid NBA, Zain ZM. Particle swarm optimization feature selection for breast cancer recurrence prediction. IEEE Access.6:29637–47 , 2018.

[10] Austria, Yolanda & Goh, Marie & Jr, Lorenzo & Lalata, Jay-Ar & Goh, Joselito & Vicente, Heintjie.Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset. International journal of simulation: systems, science & technology.10.5013/IJSSST.a.20.S2.23 , 2019 .
[11] M. Amine Naji, S. El Filali, K. Aarika, EL Habib, RAit Abdelouhahid, Debauche, Machine Learning Algorithms for Breast Cancer Prediction And Diagnosis 2021.07.062.
[12] Banu AB, Subramanian PT. Comparison of Bayes classifiers for breast cancer classification. Asian Pac J Cancer Prev (APJCP). 19(10):2917–20, 2018.
[13] Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. J Algorithms Comput Technol. 12(2):119–26 ,2018.
[14] Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Comput Appl. 23(7–8):2387–403, 2012.
[15] Tsehay Admassu Assegie , R. Lakshmi Tulasi , N. Komal Kumar.Breast cancer prediction model with decision tree and adaptive boosting.Vol. 10, No. 1, pp. 184-19, March 2021.
[16] Md. Samiul Islam, Md. Ashikuzzaman,Joy Mojumdar..Breast Cancer Detection using Machine Learning Techniques.Vol 184– No.39, 2022.
[17] Manav Mangukiya, Anuj Vaghani, Meet Savani . Breast Cancer Detection with Machine Learning 2321-9653, 2022.
[18] Shawni Dutta and Samir Kumar Bandyopadhyay,.Early Breast Cancer Prediction using Artificial Intelligence Methods.JERR.58329,2020.
[19] Malik, Shubham & Harode, Rohan & Singh, Akash. XGBoost: A Deep Dive into Boosting (Introduction Documentation). 10.13140/RG.2.2.15243.64803 , 2020..
[20] Pulung Hendro Prastyo, GedeYudi Paramartha, Michael S. Moses Pakpahan, Igi Ardiyanto.Predicting Breast Cancer: A Comparative Analysis of Machine Learning Algorithms. Volume 3, April 2020.
[21] Tharwat, Alaa. AdaBoost classifier: an overview. 10.13140/RG.2.2.19929.01122 , 2018 .
[22] Piegorsch, Walter. Confusion Matrix. 10.1002/9781118445112.stat08244 , 2020

## Authors Profile

**Md. Mijanur Rahman** is currently an assistant professor in the department of Computer Science and Engineering and Director, of the Central for Artificial Intelligence and Robotics at Southeast University. His research interests are deep learning, machine learning, medical image processing, data science and recommendation system, and blockchain. Rahman has published in conferences and journals with more than 20 pepper-reviewed journals.

**Zannatul Ferdousi** received his Bachelor's in Computer Science and Engineering at Southeast University in 2023. Her research interests are Artificial Intelligence, Deep Learning, Machine Learning, Computer Vision, and Big Data Analytics

**Puja Saha** received his Bachelor's in Computer Science and Engineering at Southeast University in 2023. Her research interests are Artificial Intelligence, Machine Learning, and Computer Vision.

**Renesha Amin Mayuri** received a Bachelor's Degree in Computer Science and Engineering at Southeast University in 2023. Her areas of interest are artificial intelligence, database systems, networking, and machine learning.