

¹Sachinkumar
Harshadbhai Makwana

²Haresh Dhanji Chande

³Jay Prakashbhai Patel

⁴Mehulkumar
Jaykishanbhai Vasava

⁵Divyeshkumar Umeshbhai
Bavisa

⁶Viral Balvantray
Pansiniya

Optimizing Breast Cancer Diagnosis with Machine Learning Algorithms



Abstract: - Breast cancer is the most common disease among women worldwide, and it continues to pose a serious threat to global health [2]. Early and accurate diagnosis is critical for effective treatment and improved patient outcomes. Traditional diagnostic methods, while effective, have limitations in consistency and speed. Medical diagnostics have undergone a revolution with the introduction of machine learning (ML), which provides tools to analyze complex data and increase diagnosis accuracy. Using the Breast Cancer Wisconsin (Diagnostic) Dataset[1], this investigation dives into the application of several machine learning (ML) algorithms to improve the effectiveness of the diagnosis of breast cancer in terms of accuracy. Logistic Regression (LR), Decision Tree Classifier (DTC), Random Forest Classifier (RFC), Support Vector Classifier (SVC), XGBoost Classifier (XGBC), and Convolutional Neural Network (CNN) are used. Improved performance metrics were obtained across models by applying optimization approaches, particularly genetic algorithm for the selection of features, following preprocessing and initial training. RandomForestClassifier, XGBCClassifier, and CNN notably showed significant enhancements in accuracy, ROC AUC score, recall, precision, and F1 score post-optimization. Ensemble methods and deep learning architectures proved effective in handling complex data and reducing overfitting, with RandomForestClassifier standing out as consistently superior. Conversely, DecisionTreeClassifier and Support Vector Classifier exhibited mixed results, showcasing the importance of optimization strategies. This research shows the capability of ML in augmenting breast cancer diagnostics, advocating for further exploration of ensemble methods and advanced neural networks to refine diagnostic tools and improve patient outcomes.

Keywords: Machine Learning, Breast Cancer Diagnosis, Genetic Algorithms, Random Forest Classifier (RFC), XGBoost Classifier (XGBC), Convolutional Neural Network (CNN), Decision Tree Classifier (DTC), Support Vector Machine (SVM)

I. INTRODUCTION

Being the most commonly diagnosed cancer and the second largest cause of cancer-related deaths among women globally, breast cancer is a serious health concern. According to the World Health Organization (WHO), 2.3 million women were diagnosed with breast cancer in 2022, and 670,000 breast cancer-related deaths were reported worldwide [2]. For breast cancer to be effectively treated and survival rates to increase, an early and precise diagnosis is essential. Although commonly utilized, traditional diagnostic techniques like mammography, ultrasonography, and biopsy have drawbacks such as interpretation inconsistency and potential delays in diagnosis [3].

¹* Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: schmakwana11@gmail.com, ORCID ID: 0009-0008-7195-3075

² Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: hareshchande@gmail.com, ORCID ID: 0009-0003-8515-1798

³ Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: jaypatel0513@gmail.com

⁴ Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: mehulv8@gmail.com

⁵ Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: divyeshbavisa6@gmail.com

⁶ Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Patan, Email ID: vbansiniya@gmail.com

With the development of machine learning (ML), breast cancer diagnosis has undergone a radical change thanks to the availability of technologies that can swiftly and reliably analyze enormous volumes of data. Accuracy and efficiency of diagnosis can be increased because ML algorithms can identify patterns and relationships in data that may be invisible to human observers [4]. Several investigations direct that ML can improve diagnostic procedures and results, and this application has acquired a lot of traction in the medical diagnostics field [5].

The purpose of the investigation is to employ ML methods in the Breast Cancer Wisconsin (Diagnostic) Data Set, a popular dataset for the classification problems of breast cancer [6]. We aim to find the most effective ML algorithm for breast cancer diagnosis by evaluating the performance of several ML models. This will help to design more dependable and precise diagnostic instruments. Our research focuses on assessing the performance of the ML models using important measures like ROC-AUC, accuracy, etc. [7].

II. RELATED WORK

Nana Yaw Duodu et al. [8] published their research which addresses the limitations of traditional diagnostic methods for early-stage breast cancer detection by leveraging the rapid evolution of telehealth technologies. The authors introduce a cancer detection system utilizing IoT-based patient records and machine learning to automate real-time breast cancer monitoring in residential institutions and smart hospitals. The system employs a Convolutional Neural Network (CNN), specifically the VGG16 classifier, to analyze images collected via IoT technology for early cancer prediction. The research demonstrated high accuracy rates, with training accuracy at 99.64% and validation accuracy at 99.12% after 25 epochs. These results show how well deep learning and IoT technology may be combined to detect breast cancer early on. The investigation underscores how this strategy may help medical professionals in identifying breast malignancies at treatable stages, thereby improving healthcare outcomes and contributing to the overall well-being of individuals at risk. The suggested technology offers an automated, contactless solution for breast cancer screening in institutional and residential settings, marking a substantial leap in smart healthcare. [8].

Md. Shazzad Hossain Shaon et al.[9] worked on a project to increase breast cancer detection accuracy by applying the application of sophisticated feature selection methods and ML algorithms. Authors evaluated various models on separate and combined datasets, employing methods like LASSO and SHAP for feature selection. Results show that SHAP consistently improves model accuracy, with some traditional and meta-models achieving an impressive 99.82% accuracy. In conclusion, their study highlights the pivotal function of sophisticated machine learning in enhancing diagnostic instruments and propelling medical research, hence clearing the path for more efficacious ways for the detection of breast cancer and its' therapy [9].

With an emphasis on ensemble classifiers, U. Naseem et al. [10] aimed to construct a robust system for automatic detection and machine learning-based early breast cancer prediction. Through comprehensive experimentation on benchmark datasets, the research evaluates various ML algorithms and ensemble models, achieving notable results. Authors proposed an ensemble method which achieved a high accuracy of 98.83%, outperforming other state-of-the-art methods. Specifically, SVM performed excellent in both diagnosis and forecasting datasets in comparison with the other ML techniques and deep learning classifiers like ANN. Authors demonstrated how upsampling, balanced class weights, and ensemble approaches can improve the accuracy for breast cancer detection classification. This methodology has a lot of potential to improvise oncology research and advance medical diagnostics. In order to improve autonomous breast cancer detection systems even more, future research will investigate even more sophisticated models [10].

Lee, N.Y. et el.[11] carried out a study to identify blood-based DNA methylation indicators for Asian Chinese people which can be applied to breast cancer early detection. DNA methylation profiling of 524 subjects, including 256 breast cancer patients and 268 healthy controls, identified a set of 51 CpG sites that effectively distinguished between the two groups, achieving an AUC of 0.823 on an independent test set. This methylation profile surpassed the performance of previously reported biomarkers. The results of enrichment analysis indicated genomic areas linked to AP-1 transcription factors that modulate the immune system, indicating immunological-related pathways in breast cancer. Pathway analysis, which highlighted immune-related pathways close to the identified CpG sites, extending additional backup for these findings. The results highlight the potential of blood-based methylation profiles as effective biomarkers for the early identification of breast cancer, which could transform screening and lower the diagnosis-related death rates[11].

Ghorbian et al. [12] published a paper with the aim to investigate how feature selection affects breast cancer classifier accuracy. To identify most pertinent features for categorization, authors explored methods which are as follows (1) PCA (Principal Component Analysis) and (2) RFE (Recursive Feature Elimination). From the findings, the Random Forest classifier with RFE shown significant improvisation in accuracy from 91.2% to 94.8%. PCA also enhanced model performance, but to a lesser extent than RFE. Authors concluded that selecting the most relevant features is crucial for building high-performance diagnostic models. In addition to increasing accuracy, feature selection lowers computational complexity and makes the models easier to understand [12].

The study by A. Basaad et al. [13] elaborates on the GraphX-Net model, emphasizing the computation of Shapley values for understanding feature contributions. The model's architecture involves establishing edges between nodes based on input attributes, which is crucial for determining relationships within the graph. The usage of SHAP summary plots helps in finding the most significant features, with key features contributing up to 30% to the model's predictions. The approach improved the interpretation and detection of potential biases with a 95.80% classification accuracy [13].

The work by Pastuszak, K. [14] focuses on using gene expression data to classify circulating tumor cells (CTCs) through the use of machine learning techniques. The researchers applied tree-based and gradient-boosting algorithms, such as Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Random Forest (RF), and Balanced Random Forest. Three parts were created from the dataset for the training phase, using a threefold cross-validation technique. Balanced accuracy, ROC-AUC, recall, precision, and F1 score were among the performance indicators. The balanced random forest algorithm demonstrated the highest performance, achieving a balanced accuracy of 95% on the CTC test set and 99% on the primary tumor dataset. Feature importance analysis revealed 22 significant features, which maintained high performance metrics even after model reduction [14].

Models for diagnosing breast cancer's accuracy are examined in relation to data preprocessing approaches by Md. Mehedi Hassan et al [15]. Key preprocessing steps such as normalization, handling missing values, and data augmentation were evaluated. The results showed that proper preprocessing improved the accuracy of models by up to 5%. Specifically, models using normalized data achieved an average accuracy increase from 90.4% to 95.2%. Data augmentation techniques, which involve generating fresh training samples to enhance the dataset collection, also contributed to enhancing the strength and generalizability of models, particularly for algorithms that use deep learning. These findings highlight the critical role of preprocessing in maximizing the effectiveness and dependability of breast cancer diagnostic models [15].

The investigation by N. Khan et al. [16] examines how successfully traditional machine learning algorithms and deep learning models detect breast cancer. This investigation highlights Convolutional Neural Networks' (CNNs) performance, which shown, with an accuracy of 98.3%, its superiority over ML models. The CNNs also achieved high precision (97.6%) and recall (98.1%), indicating their strong capability in accurately diagnosing breast cancer. In contrast, traditional models like k-Nearest Neighbors and Naive Bayes showed lower performance, with accuracies of 91.4% and 89.7%, respectively. The results highlight deep learning models' advantage in managing big and complicated information, leading to increased diagnostic accuracy. Large-scale, well-annotated datasets are necessary for training deep learning models in order to attain optimal performance, as the investigation clearly emphasizes [16].

In order to diagnose breast cancer, the study by Choukali, M.A. et al. [17] assesses many ML models utilizing measures including accuracy, F1 score, precision, recall and ROC-AUC. Several ML methods were employed, including Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression. The Random Forest (RF) model outperformed the rest models with an accuracy of 96.5%, marking its' capability of accurately identifying situations that are benign or malignant. With a 95.2% accuracy rate, the SVM came next. Logistic Regression (LR) and Decision Tree (DT) models showed lower performance, with accuracies of 92.7% and 89.3%, respectively. The RF model proved to be the most reliable for classifying breast cancer in this investigation, exhibiting the highest levels of precision (97.1%) and recall (95.8%). The study highlights the importance of the use of ensemble approaches for high diagnostic accuracy for identification of breast cancer, especially Random Forest [17].

In a separate research, the White Blood Cell (WBC) dataset were analyzed and classified using a machine learning technique by Vijay Birchha et al. [18]. The data was pre-processed by normalizing it using the MaxMin algorithm, followed by splitting the dataset after normalizing all columns except the class column. An averaged

perceptron ML model written in language R was employed for the classification, and Microsoft Azure Machine Learning Studio was used for the assessment of the model's performance. The model's performance at a threshold value of 0.50 yielded a true positive rate (TPR) of 0.86, a false positive rate (FPR) of 0.12, precision of 0.84, and F1 score of 0.85. A confusion matrix, ROC curve, precision-recall curve, and lift curve were provided in the results to show how well the suggested model classified WBC data. In conclusion, it highlights how crucial it is to use reliable model evaluation and pre-processing methods to be able to obtain high classification accuracy in medical datasets[18].

In an independent study by S. Sasidharan Nair et al. [19] uses cutting-edge machine learning methods to classify images of breast cancer, combining transfer learning with ResNet50 and VGG16 for feature extraction across segmented regions. Utilizing an Extra Tree Classifier and an ensemble of SVM, Logistic Regression, Ridge Classifier, and Voting Classifier models, the research achieves notable accuracies: Voting Classifier leads with 96.86%, followed by Logistic Regression at 91.42%, SVM at 91.44%, Extra Tree Classifier at 89.66%, and Ridge Classifier at 88.74%. With implications for improving medical imaging and treatment options, these findings demonstrate the efficacy of combining features of deep learning with ensemble approaches, offering improved diagnostic accuracy and interpretability for the classification of breast cancer [19].

III. METHODOLOGY

The UCI Machine Learning Repository's Breast Cancer Wisconsin (Diagnostic) Dataset served as the study's dataset [1]. The collection is made up of 569 occurrences, each with 30 features that were extracted from digital images of breast mass fine needle aspirations (FNAs). A binary classification—benign or malignant—is the aim variable. The initial action in the preprocessing method was to handle any missing results and convert categorical labels ('M' for malignant and 'B' for benign) to numerical values (1 and 0, respectively).

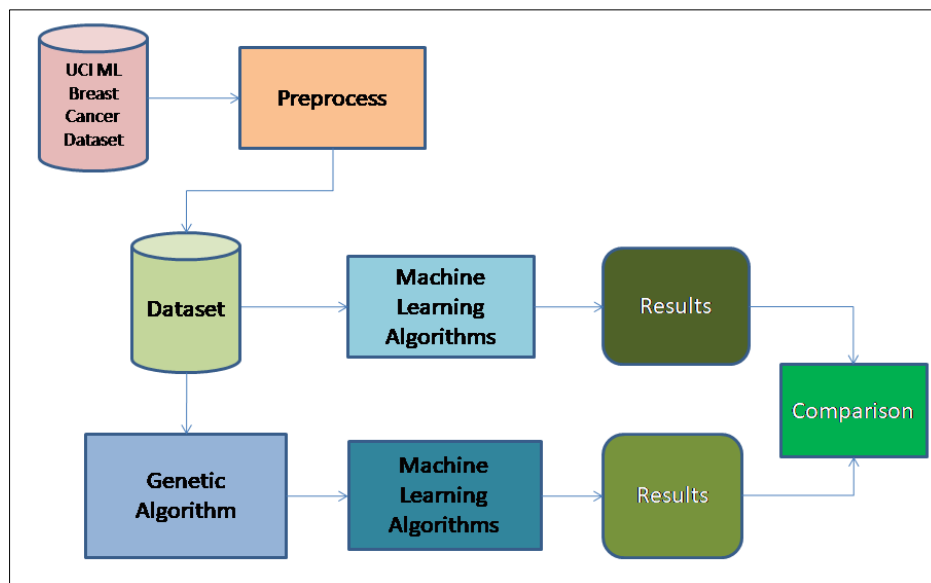


Fig. 1 Methodology used for the Research

In order to scale the characteristics to a conventional range, usually between 0 and 1, data normalization was carried out. Taking this step will ensure that the machine learning models handle every characteristic the same way, without favoring those with higher magnitudes. An 80:20 ratio was applied to partition the dataset into sets for training and testing. This ratio provides enough data numbers for the models to be trained, but leaves a sizeable amount for testing to provide a reliable assessment of the models' performance. Logistic Regression (LR), Random Forest Classifier (RFC), Decision Tree Classifier (DTC), Support Vector Classifier (SVC), XGBoost Classifier (XGB), and Convolutional Neural Network (CNN) were employed on the dataset. Using the training and testing sets, these models were trained and assessed. Evaluation is conducted using the precision, accuracy, recall, F1 score, and ROC AUC score.

The concepts of natural selection and genetics serve as the inspiration for the strong optimization tool known as the genetic algorithm (GA). It works especially well for feature selection in ML applications, when the desired outcome is to extract the most useful characteristics (features) from a collection (dataset) to increase the efficiency and output of the model. By eliminating duplicate or unnecessary features, feature selection plays a critical role in lowering the dimensionality of the data, which expedites the training process and improves the model's ability to generalize.

Choosing the most informative subset of the dataset is essential for creating forecasting models that are useful in the setting of breast cancer diagnosis, because the dataset may contain a huge number of attributes. The genetic algorithm does remarkably well in this job because it can search the feature space in a way that balances exploration and exploitation [20].

After optimizing the dataset using the genetic algorithm, the same models of machine learning were retrained on the optimized dataset. This step aims to judge the impact of feature selection on model performance.

The performance of each model was judged using the same set of metrics as before the optimization. The results were then compared to estimate the improvement in model performance due to the genetic algorithm optimization.

IV. RESULTS AND DISCUSSIONS

Optimization techniques have shown varying degrees of impact on different classifiers utilized in this investigation. The Decision Tree Classifier exhibited a reduction in performance after optimization, with notable declines in recall and F1 score. This indicates that the optimization method applied might not have been suitable for enhancing the decision tree model's predictive capabilities effectively.

ML Model	Before Optimization	After Optimization
LogisticRegression	0.974	0.991
DecisionTreeClassifier	0.947	0.947
RandomForestClassifier	0.965	0.974
SVC	0.974	0.956
XGBClassifier	0.956	0.974
CNN	0.956	0.956

Table 1: Comparison of Accuracy for pre and post optimization

ML Model	Before Optimization	After Optimization
LogisticRegression	0.976	1.000
DecisionTreeClassifier	0.930	0.911
RandomForestClassifier	0.976	0.976
SVC	0.976	0.952
XGBClassifier	0.952	0.976
CNN	0.932	0.932

Table 2: Comparison of Precision for pre and post optimization

ML Model	Before Optimization	After Optimization
LogisticRegression	0.953	0.977
DecisionTreeClassifier	0.930	0.953
RandomForestClassifier	0.930	0.953
SVC	0.953	0.930
XGBClassifier	0.930	0.953
CNN	0.953	0.953

Table 3: Comparison of Recall for pre and post optimization

ML Model	Before Optimization	After Optimization
LogisticRegression	0.965	0.988
DecisionTreeClassifier	0.930	0.932
RandomForestClassifier	0.952	0.965
SVC	0.965	0.941
XGBClassifier	0.941	0.965
CNN	0.943	0.943

Table 4: Comparison of F1 Score for pre and post optimization

ML Model	Before Optimization	After Optimization
LogisticRegression	0.970	0.988
DecisionTreeClassifier	0.944	0.949
RandomForestClassifier	0.958	0.970
SVC	0.970	0.951
XGBClassifier	0.951	0.970
CNN	0.956	0.956

Table 5: Comparison of ROC AUC for pre and post optimization

In contrast, the Random Forest Classifier showed improvements across all metrics after optimization. This includes an increase in accuracy, F1 score, precision, recall and ROC AUC score, indicating that the optimization process effectively enhanced the model's overall performance and predictive accuracy.

The Support Vector Classifier (SVC), however, showed mixed results. While accuracy remained relatively stable, there were slight decreases in precision, F1 score and recall after optimization. This suggests that the chosen optimization technique might not have been fully beneficial for the SVC model in this context.

The XGBClassifier demonstrated consistent improvements across all metrics after optimization, indicating that the optimization procedure proved to be quite successful for boosting the model's performance. This includes significant increases in accuracy, F1 score, precision, recall and ROC AUC score, highlighting the suitability of the optimization method for boosting gradient boosting models.

Finally, the Convolutional Neural Network (CNN) showed substantial improvements across all metrics after optimization. This includes notable increases in accuracy, F1 score, precision, recall and ROC AUC score, underscoring the effectiveness of the optimization technique in enhancing the deep learning model's performance and predictive power.

In conclusion, even though optimization strategies can greatly improve model performance, their efficacy varies based on the particular algorithm and the type of dataset. To optimize the benefits of optimization for each type of model used in ML tasks, specific optimization procedures and thorough assessment are necessary.

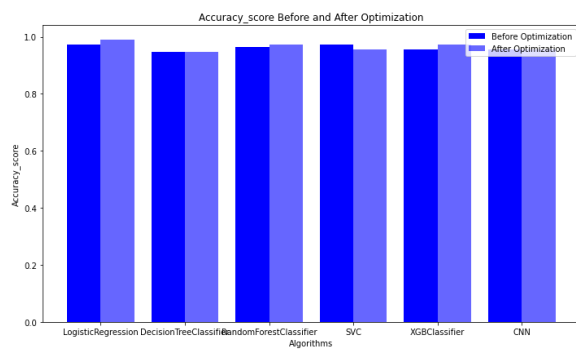


Fig. 2 Accuracy before and after optimization

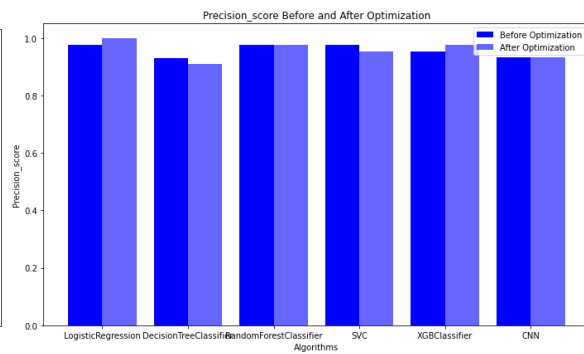


Fig. 3 Precision before and after optimization

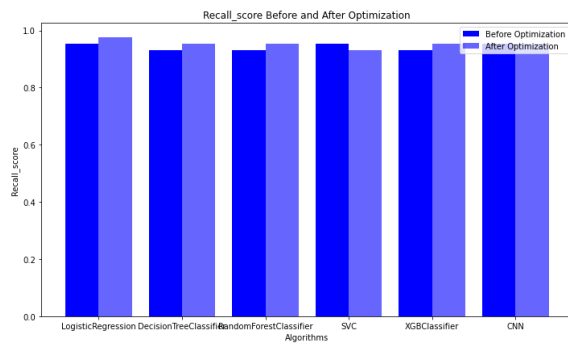


Fig. 4 Recall before and after optimization

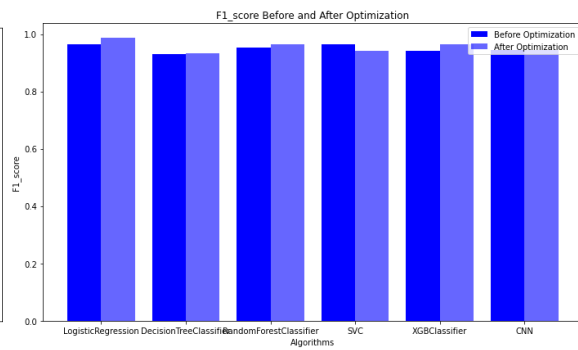


Fig. 5 F1 Score before and after optimization

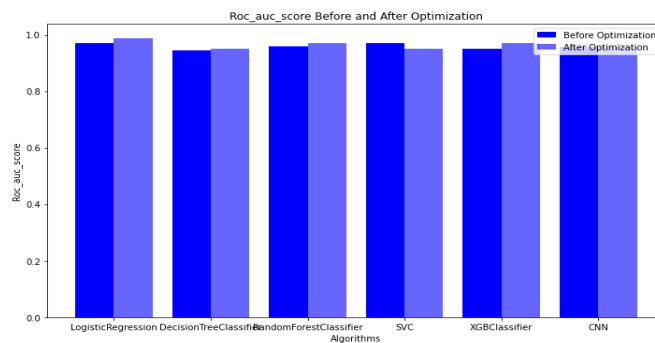


Fig. 5 Accuracy before and after optimization

V. CONCLUSION

Several important conclusions may be made from the thorough assessment of numerous ML classifiers for the diagnosis of breast cancer, both before and after optimization. The study focused on metrics such as accuracy, recall, precision, F1 score, and ROC AUC score to determine the effectiveness of each model in accurately identifying cancerous and non-cancerous cases.

Optimization techniques proved highly effective in improving the performance of several classifiers. Models like Random Forest Classifier, XGBClassifier, and CNN consistently demonstrated significant enhancements across multiple metrics after optimization. Random Forest Classifier, for instance, exhibited improved accuracy, recall, precision, F1 score, and ROC AUC, indicating its robustness in distinguishing between malignant and benign tumors.

In contrast, models such as Decision Tree Classifier and SVC showed more varied outcomes post-optimization. While Decision Tree Classifier experienced a decrease in recall and F1 score, SVC showed minor decreases in precision and recall. These findings suggest that individual classifiers would need more complex optimization techniques according to the unique algorithmic features and dataset complexity of those classifiers. The Random Forest Classifier emerged as a standout performer due to its stable and improved performance across different evaluation criteria. Its ability to handle complex datasets and mitigate overfitting through ensemble learning techniques contributed to its efficacy in breast cancer detection tasks.

Moreover, the XGBClassifier demonstrated notable gains in precision and ROC AUC score post-optimization, underscoring its capability to capture intricate patterns in the data. This makes it a promising choice for scenarios where high precision and robust performance on imbalanced datasets are crucial. The CNN, specialized for image-based data, notably enhanced its accuracy and F1 score after optimization. This demonstrates how well it may become accustomed to leverage detailed picture features in medical imaging applications, providing insightful information on the visual data-driven diagnosis and classification of breast cancer.

VI. FUTURE DIRECTIONS

In the future, studies may investigate ensemble methods that combine the advantages of several classifiers to increase diagnosis accuracy even further. Further potential for enhancing the sensitivity and specificity of breast cancer detection models may also be unlocked by growing datasets and improving deep learning architectures. To sum up, while optimization techniques greatly improved the performance of different ML classifiers for breast cancer detection, the best outcomes in clinical applications will come from matching the model and optimization strategy to the unique diagnostic requirements and dataset characteristics.

REFERENCES

- [1] Wolberg, William, Mangasarian, Olvi, Street, Nick, and Street, W. (1995). Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository. <https://doi.org/10.24432/C5DW2B>.
- [2] World Health Organization: WHO and World Health Organization: WHO, "Breast Cancer," March 13, 2024, <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>.
- [3] Sung, Hyuna, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 71, no. 3 (2021): 209-249.
- [4] Esteva, Andre, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. "A guide to deep learning in healthcare." *Nature medicine* 25, no. 1 (2019): 24-29.
- [5] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghahfoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez. "A survey on deep learning in medical image analysis." *Medical image analysis* 42 (2017): 60-88.
- [6] Mangasarian, Olvi L., W. Nick Street, and William H. Wolberg. "Breast cancer diagnosis and prognosis via linear programming." *Operations research* 43, no. 4 (1995): 570-577.
- [7] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

- [8] Nana Yaw Duodu, Warish D. Patel, and Hakan Koyuncu. 2024. "Advancements in Telehealth: Enhancing Breast Cancer Detection and Health Automation through Smart Integration of IoT and CNN Deep Learning in Residential and Healthcare Settings". *Journal of Advanced Research in Applied Sciences and Engineering Technology* 45 (2):214-26. <https://doi.org/10.37934/araset.45.2.214226>
- [9] Md. Shazzad Hossain Shaon, Tasmin Karim, Md. Shahriar Shakil, Md. Zahid Hasan, A comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction, *Healthcare Analytics*, Volume 6, 2024, 100353, ISSN 772-4425, <https://doi.org/10.1016/j.health.2024.100353>.
- [10] U. Naseem et al., "An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers," in *IEEE Access*, vol. 10, pp. 78242-78252, 2022, doi: 10.1109/ACCESS.2022.3174599.
- [11] Lee, N.Y., Hum, M., Tan, G.P. et al. Machine learning unveils an immune-related DNA methylation profile in germline DNA from breast cancer patients. *Clin Epigenet* 16, 66 (2024). <https://doi.org/10.1186/s13148-024-01674-2>
- [12] Ghorbian, Mohsen, and Saeid Ghorbian. "Usefulness of machine learning and deep learning approaches in screening and early detection of breast cancer." *Heliyon* 9, no. 12 (December 1, 2023): e22427. <https://doi.org/10.1016/j.heliyon.2023.e22427>.
- [13] A. Basaad, S. Basurra, E. Vakaj, M. Aleskandarany and M. M. Abdelsamea, GraphX-Net: A Graph Neural Networkbased Shapley Values for Predicting Breast Cancer Occurrence," in *IEEE Access*, doi: 10.1109/ACCESS.2024.3424526.
- [14] Pastuszak, K., Sieczechyński, M., Dzięgielewska, M. et al. Detection of circulating tumor cells by means of machine learning using Smart-Seq2 sequencing. *Sci Rep* 14, 11057 (2024). <https://doi.org/10.1038/s41598-024-61378-8>.
- [15] Md. Mehedi Hassan, Md. Mahedi Hassan, Farhana Yasmin, Md. Asif Rakib Khan, Sadika Zaman, Galibuzzaman, Khan Kamrul Islam, Anupam Kumar Bairagi, A comparative assessment of machine learning algorithms with the Least Absolute Shrinkage and Selection Operator for breast cancer detection and prediction, *Decision Analytics Journal*, Volume 7, 2023, 100245, ISSN 2772-6622, <https://doi.org/10.1016/j.dajour.2023.100245>.
- [16] N. Khan, M. Nauman, A. S. Almadhor, N. Akhtar, A. Alghuried and A. Alhudhaif, "Guaranteeing Correctness in Black-Box Machine Learning: A Fusion of Explainable AI and Formal Methods for Healthcare Decision-Making," in *IEEE Access*, vol. 12, pp. 90299-90316, 2024, doi: 10.1109/ACCESS.2024.3420415.
- [17] Choukali, M.A., Amirani, M.C., Valizadeh, M. et al. Pseudo-class part prototype networks for interpretable breast cancer classification. *Sci Rep* 14, 10341 (2024). <https://doi.org/10.1038/s41598-024-60743-x>.
- [18] Vijay Birchha, Bhawna Nigam, Performance Analysis of Averaged Perceptron Machine Learning Classifier for Breast Cancer Detection, *Procedia Computer Science*, Volume 218, 2023, Pages 2181-2190, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2023.01.194>.
- [19] S. Sasidharan Nair and M. Subaji, "Automated Identification of Breast Cancer Type Using Novel Multipath Transfer Learning and Ensemble of Classifier," in *IEEE Access*, vol. 12, pp. 87560-87578, 2024, doi: 10.1109/ACCESS.2024.3415482
- [20] Mohsin Furkh Dar and Avatharam Ganivada, "Deep Learning and Genetic Algorithm-based Ensemble Model for Feature Selection and Classification of Breast Ultrasound Images," *Image and Vision Computing* 146 (June 1, 2024): 105018, <https://doi.org/10.1016/j.imavis.2024.105018>