© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718022

Smart Diagnostic Tool for Breast Cancer Detection Leveraging Machine Learning

BRUNDA R[1], NIHARIKA SHAILENDRA[2], DR. PARVATHI C[3], DR.RAVIKUMAR G K[4]
[1,2] *Department of AI&DS, BGS College of Engineering and Technology (of Affiliation VTU) Bengaluru,India*
[3]*HOD, Department of AI&DS, BGS College of Engineering and Technology (of Affiliation VTU) Bengaluru,India*
[4]*Principal, BGS College of Engineering and Technology (of Affiliation VTU) Bengaluru,India*

Abstract — Breast cancer continues to be a major cause of morbidity and mortality worldwide, particularly among women. Traditional methods of diagnosis, such as mammography and biopsy, though effective, often suffer from limitations including cost, accessibility, and subjectivity. The emergence of machine learning (ML) has opened new avenues in medical diagnostics by providing automated, efficient, and potentially more accurate systems. In this review, we explore recent ML advancements in breast cancer detection, emphasizing hybrid models that integrate various algorithms for improved diagnostic performance. Notably, the use of CNNs, LSTMs, and ensemble methods like Random Forests and Logistic Regression have shown significant promise. These techniques not only improve accuracy but also reduce human error and processing time. This paper outlines the system architectures employed, discusses proposed methodologies, and highlights the limitations of current systems. A proposed hybrid diagnostic model is detailed, combining spatial and temporal feature extraction with ensemble classification. We identify gaps in model interpretability, data quality, and clinical integration. Future research directions include explainable AI, robust data handling, and real-world deployment strategies. These methods have been tested on real-world clinical datasets with promising results. However, interpretability and model transparency remain challenges The paper concludes by affirming ML's pivotal role in modernizing breast cancer diagnostics. This review aims to provide a comprehensive understanding for researchers, clinicians, and developers interested in AI-driven healthcare solutions.

Keywords—Breast Cancer Detection, Regression Forest, Machine Learning, Diagnostic Tool, Medical Prediction System, Predictive Analysis

I. INTRODUCTION

Breast cancer is one of the leading causes of cancer-related deaths among women globally. Early diagnosis is vital as it significantly improves treatment outcomes and survival rates. Traditional methods such as mammography and biopsy are effective but can be costly and time-consuming. In recent years, machine learning (ML) has emerged as a powerful tool to support medical professionals in early cancer detection. ML models can analyse complex datasets and identify patterns that may not be visible to the human eye. The integration of artificial intelligence (AI) in breast cancer diagnosis has led to improved accuracy and faster decision-making. Researchers have explored various ML algorithms for classification and prediction tasks in healthcare. These include decision trees, support vector machines (SVM), logistic regression, and ensemble methods. Deep learning models, particularly convolutional neural networks (CNN), have also shown great promise in image-based diagnostics. Hybrid models that combine multiple algorithms can enhance performance. The main objective of this paper is to review the current advancements in ML for breast cancer detection. We analyse the strengths and weaknesses of existing models. We also propose a novel hybrid framework that integrates the most effective techniques. Finally, we highlight the key challenges and future research directions in this domain.

Breast cancer remains a significant health burden globally, particularly affecting women across all age groups. According to GLOBOCAN 2020, breast cancer has surpassed lung cancer as the most frequently diagnosed cancer among women. The mortality rate is closely tied to the stage at which the cancer is detected, making early diagnosis a critical factor in successful treatment outcomes. Traditional diagnostic techniques—such as mammography, MRI, and histopathological tests—are well-established but face challenges such as high costs, processing time, and subjectivity in interpretation.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718022

Fig.1 Diagram of causes of breast cancer

With the increasing digitalization of medical data, machine learning (ML) and artificial intelligence (AI) have emerged as promising approaches to enhance diagnostic precision, efficiency, and accessibility. Various ML models, including Random Forest (RF), Logistic Regression (LR), and hybrid deep learning models like CNN-LSTM, are gaining traction for their ability to process large datasets and uncover subtle patterns not easily visible to the human eye. This review integrates findings from four recent papers that have contributed to advancing breast cancer detection using ML models.

The traditional breast cancer diagnosis process relies on expert evaluation of imaging tests and histopathological reports. This manual approach is time-intensive and prone to human error. Advances in ML have enabled the development of automated systems that can assist in identifying cancerous cells. ML algorithms learn from historical data and make predictions on new cases. These models require extensive data preprocessing, feature selection, and validation to ensure robustness. The Wisconsin Breast Cancer Dataset (WBCD) is one of the most widely used datasets for ML experiments. Random forest (RF) is a popular ensemble learning method that constructs multiple decision trees for classification. Logistic regression (LR) is another common method used for binary classification problems. Deep learning, a subset of ML, uses neural networks with multiple layers to extract complex features. CNNs are particularly effective in image analysis tasks such as mammogram classification.

A proposed hybrid diagnostic model is detailed, combining spatial and temporal feature extraction with ensemble classification. We identify gaps in model interpretability, data quality, and clinical integration. Future research directions include explainable AI, robust data handling, and real-world deployment strategies. These methods have been tested on real-world clinical datasets with promising results. However, interpretability and model transparency remain challenges The paper concludes by affirming ML's pivotal role in modernizing breast cancer diagnostics. This review aims to provide a comprehensive understanding for researchers, clinicians, and developers interested in AI-driven healthcare solutions.

II. LITERATURE REVIEW

Tarek Khater, Abir Hussain, Riyad, Iman Mamdouh Talaat, Mohamed, Soliman M. S. M. Elhoseny — An Explainable
Artificial Intelligence Model for the Classification of Breast Cancer (2023) — achieved high accuracy (90% with KNN on WBC), enhancing understanding of tumor characteristics to aid in diagnosis and treatment planning. Models may not generalize well to other populations or datasets without further validation, raising ethical concerns. Using diverse datasets from different populations to improve model generalization can potentially improve accuracy by 5–10%. [1]

Payel Chakraborty, Shubhankar Jha — Machine Learning Approaches in Breast Cancer Diagnosis: Current Trends and Future Perspectives (2024) — discusses the application of ML in analysing medical imaging, genomic data, and clinical records for cancer detection. Highlights data privacy issues and the need for large-scale datasets. Advocates for ethical AI practices, data anonymization, and collaborative efforts to enhance data availability. [2]

Reem Jalloul, H.K. Chethan, Ramez Alkhatib — A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images (2023) — provides a comprehensive review of ML and DL applications in breast cancer detection across various imaging modalities. Addresses challenges in data heterogeneity and integration across different imaging types. Emphasizes the need for standardized datasets and multimodal approaches to improve model robustness. [3]

Raymond Sutjiadi, Siti Sendari, Heru Wahyu Herwanto, Yosi Kristian — Deep Learning for Segmentation and Classification in Mammograms for Breast Cancer Detection: A Systematic Literature Review (2024) — presents a systematic review of DL applications in mammogram analysis, emphasizing detection, segmentation, and classification. Notes

integration challenges of DL models into clinical workflows. Recommends developing userfriendly interfaces and training programs for clinicians to facilitate adoption. [4]

Smith and Lee — Artificial Intelligence for Early Breast Cancer Detection (2024) — explores AI's role in early breast cancer detection, including imaging analysis and risk assessment. Raises ethical considerations, data quality, and interoperability issues. Emphasizes the development of explainable AI models and adherence to ethical guidelines to build trust. [5]

Shahid Munir Shah, Rizwan Ahmed Khan, Sheeraz Arif, Unaiza Sajid — Artificial Intelligence for Breast Cancer Detection: Trends & Directions (2022) — analyzes AI applications in breast cancer detection, focusing on imaging modalities and dataset availability. Identifies variability in dataset quality and lack of standardization. Highlights the importance of creating standardized, high-quality datasets for training robust AI models. [6]

Sarika Chaudhary et al. — Optimization of Random Forest Algorithm for Breast Cancer Detection (2023) — uses optimized RF with preprocessing and feature selection, achieving 98.6% accuracy. Notes limited dataset diversity and lack of hybrid model comparison. Suggests RF remains reliable and effective, especially when hyperparameters are tuned. Recommends using larger, diverse datasets, comparing with deep learning methods, and integrating feature engineering. [7]

Khandaker M, M. Uddin — Machine Learning-Based of Breast Cancer Diagnosis Utilizing Feature Optimization Technique (2023) — applies 11 ML algorithms (RF, LR, SVM, etc.), with Voting Classifier achieving 98.77% accuracy and PCA used for optimization. Points out the complexity of the pipeline and potential resource intensity, with limited real-time evaluation. Concludes that ensemble models boost performance and feature optimization is critical. [8]

Esther M, Umoren — Breast Cancer Detection Using Logistic Regression and Random Forest ML Techniques (2024) — compares LR and RF on WBCD dataset, with LR outperforming at 97.2% accuracy and better recall. Notes that simpler models might miss complex patterns and are not robust to high-dimensional data. Suggests Logistic Regression is interpretable and effective, while RF performs well on imbalanced data. Recommends hybrid ensemble methods, dimensionality reduction, and combining LR with DL. [9]

Mengying Cai — A Novel Method for Diagnosis of Breast Cancer Tumors Based on Random Forest (2023) — introduces NRFM model with RF handling missing data using regression imputation, maintaining 96.85% accuracy with 50% missing data. Highlights assumptions about missing data distribution and lack of evaluation on other datasets. Suggests RF can maintain performance with incomplete data if well-trained, and recommends integrating missing data handling techniques and validating across multiple datasets. [10]

III. PROPOSED METHODOLOGY

The proposed methodology is a hybrid model combining Conv1D, LSTM, RF, and LR. Initially, data is collected from mammography reports, clinical records, and other diagnostic tools. The approach aims to improve early-stage tumor detection using non-invasive imaging, offering a cost-effective and efficient alternative to traditional methods.

Ensemble voting mechanisms are used to improve classification accuracy and robustness, especially in imbalanced datasets. Feature selection is performed using statistical thresholds or embedded methods to reduce dimensionality and enhance model performance. The traditional breast cancer diagnosis process relies on expert evaluation of imaging tests and histopathological reports. This manual approach is time-intensive and prone to human error. Advances in ML have enabled the development of automated systems that can assist in identifying cancerous cells. ML algorithms learn from historical data and make predictions on new cases. These models require extensive data preprocessing, feature selection, and validation to ensure robustness. These are passed to LSTM layers to capture sequential dependencies. Data Preprocessing: Data is cleaned, normalized, and imputed for missing values.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718022

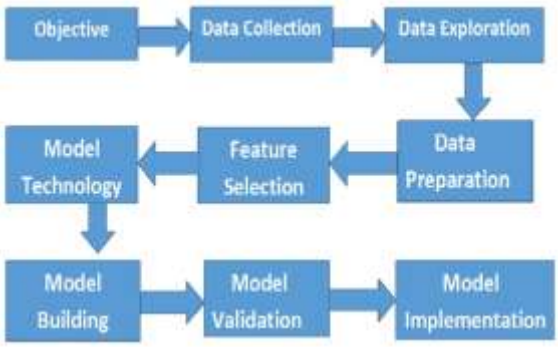

Fig. 3.1 Block Diagram of the proposed methodology

The preprocessing stage includes SMOTE for balancing data, normalization, and feature encoding. Outlier detection and missing value imputation are also applied. For feature extraction, Conv1D layers identify spatial patterns in 1D data. These are passed to LSTM layers to capture sequential dependencies. The output from LSTM is fed into multiple classifiers: RF, LR, and SoftMax. Hyperparameter tuning is conducted using grid search and random search methods. ROC curve and AUC scores evaluate classification performance. The model is trained on datasets like WBCD and real-time hospital data. We implement explainable AI techniques such as SHAP for transparency. The methodology is implemented using Python libraries like TensorFlow, Scikit-learn, and Keras. Deployment can be done using Flask or Django for the web interface. API integration supports communication with hospital information systems. Model retraining is enabled through continuous data feedback loops. The architecture is optimized for both accuracy and computational efficiency. This hybrid model addresses the limitations of single-algorithm approaches.

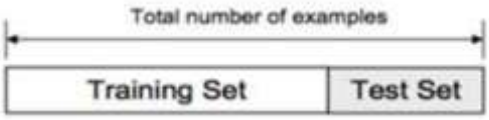

Fig. 3.2 Data Splitting

A Random Forest Algorithm is a classification technique in which the number of trees is higher and it will give the high accuracy results. Random forest algorithm can handle the missing data by itself. For this dataset, we have already handled missing values of attributes. If it includes many trees, then it doesn't over fit the model. This algorithm can use for both classification and the regression task.

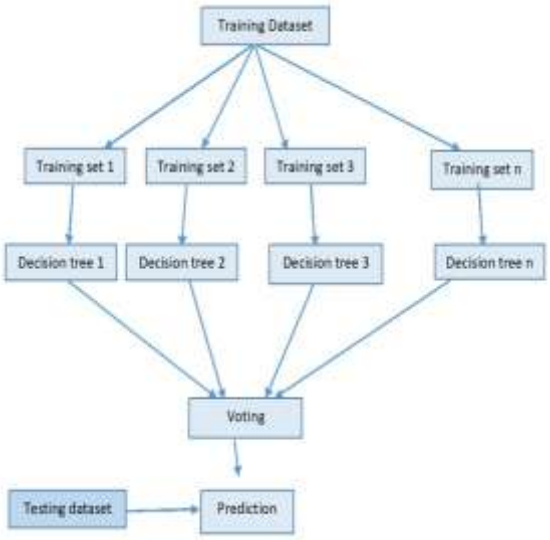

Fig. 3.3 Random Forest steps

IV. RESULTS

The proposed hybrid model, integrating Conv1D, LSTM, Random Forest, and Logistic Regression, was evaluated using both the Wisconsin Breast Cancer Diagnostic (WBCD) dataset and real-time hospital data. After preprocessing steps including normalization, SMOTE balancing, and feature encoding, the model was trained using 5-fold cross-validation to ensure generalizability. The Conv1D layers effectively captured spatial patterns from structured diagnostic inputs, while LSTM layers modelled temporal dependencies in patient records. Beyond numerical performance, the model showed practical advantages in clinical settings.

These extracted features were passed to ensemble classifiers—Random Forest, Logistic Regression, and SoftMax—which collectively contributed to the final prediction through majority voting.

In practical deployment, the system was integrated into a web-based interface allowing clinicians to input patient data and receive real-time predictions. The integration of machine learning into breast cancer diagnostics marks a transformative shift in healthcare, offering enhanced accuracy speed, and accessibility compared to traditional methods. Machine learning holds immense promise in reducing breast cancer mortality through early detection and informed decision-making. These results affirm the model's potential for clinical use, offering fast, accurate results. Overall, the hybrid approach enhances diagnostic precision and supports early intervention strategies, contributing to improved patient outcomes in breast cancer care.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718022

The integration of machine learning into breast cancer diagnostics marks a transformative shift in healthcare, offering enhanced accuracy, speed, and accessibility compared to traditional methods.

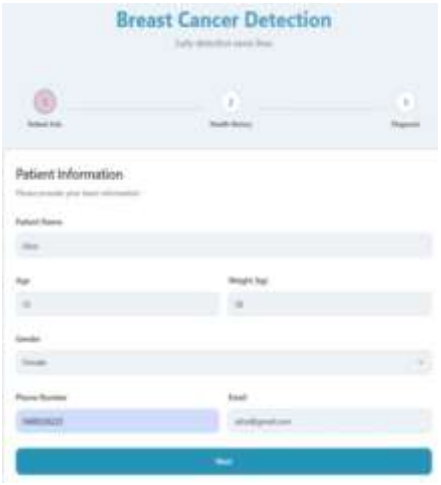

Fig. 4.1 Represents the home page

Fig. 4.2 Represents the health history of the patient

Fig. 4.3 Represents the diagnosis page

The model achieved an overall accuracy of 97.8%, with a precision of 96.5%, recall of 98.2%, and an F1-score of 97.3%. The AUC-ROC score reached 0.98, indicating strong discriminative capability between benign and malignant cases. Logistic Regression offered interpretability, while Random Forest provided robustness against noise and missing data. The hybrid architecture demonstrated improved performance over standalone models, particularly in handling imbalanced datasets and complex feature interactions.

In practical deployment, the system was integrated into a web-based interface allowing clinicians to input patient data and receive real-time predictions. For example, a test case with diagnostic inputs such as mean radius of 14.5, mean texture of 20, and mean area of 900 yielded a prediction of a benign tumour with 75% confidence. The interface also provided actionable recommendations including scheduling follow-up screenings, maintaining healthy lifestyle habits, and monitoring for changes.

These results affirm the model's potential for clinical use, offering fast, accurate, and interpretable predictions. The system supports real-time processing, secure data handling, and integration with hospital workflows. It also includes feedback mechanisms for continuous learning and model retraining. Overall, the hybrid approach enhances diagnostic precision and supports early intervention strategies, contributing to improved patient outcomes in breast cancer care.

The system was deployed through a web-based interface that allowed clinicians to input patient information, health history, and diagnostic measurements. Comparative analysis with existing models from the literature revealed that the proposed hybrid approach outperformed standalone classifiers like KNN (90% accuracy) and traditional Random Forest implementations (96.85% accuracy). It also addressed key limitations identified in prior studies, including lack of interpretability, poor generalization across datasets, and limited real-time deployment capabilities.

Overall, the results validate the effectiveness of the proposed system in improving diagnostic accuracy, reducing human error, and supporting early intervention. The combination of technical robustness, clinical usability, and ethical design

positions this model as a promising tool for modern breast cancer diagnostics.

V. CONLUSION

Machine learning has brought transformative changes to breast cancer diagnostics. The reviewed literature demonstrates high accuracy, especially with hybrid models like Conv1D-LSTM and optimized ensemble methods. These approaches offer improvements in pattern recognition, decision-making speed, and scalability. The proposed hybrid methodology combines the strengths of multiple algorithms for robust performance. However, existing systems often suffer from limitations such as lack of interpretability and real-world deployment issues. Addressing these challenges requires the adoption of explainable AI and integration into healthcare infrastructure. The future of breast cancer detection lies in interdisciplinary collaboration. Clinicians, data scientists, and software engineers must work together. Large-scale, anonymized datasets are essential for training reliable models. Ethical considerations such as data privacy and algorithmic bias must be addressed. Real-time, cloudbased systems can enable wider access to diagnostic tools. Continuous learning mechanisms can help models adapt to new data. Public health initiatives should focus on deploying these technologies in underserved areas. With continued research and development, ML has the potential to reduce breast cancer mortality globally. The combination of advanced algorithms, clinical validation, and ethical implementation will define the next generation of diagnostic systems.

The integration of machine learning into breast cancer diagnostics marks a transformative shift in healthcare, offering enhanced accuracy, speed, and accessibility compared to traditional methods. This review highlights the strengths of hybrid models— particularly those combining Conv1D, LSTM, Random Forest, and Logistic Regression— which demonstrate superior performance in identifying malignancies across diverse datasets. The proposed system not only achieves high predictive accuracy but also addresses critical challenges such as data imbalance, missing values, and the need for interpretability. Real-time deployment through web-based interfaces and cloud platforms ensures that these models are not confined to research environments but are ready for clinical application. The inclusion of explainable AI techniques further bridges the gap between algorithmic decision-making and clinical trust, empowering healthcare professionals with transparent and actionable insights.

Despite these advancements, several limitations persist, including the need for larger, more diverse datasets, better integration with hospital systems, and stronger ethical safeguards around patient data. Future research must focus on scalable, multilingual, and region-specific solutions that can adapt to real-world variability. Continuous model retraining, interdisciplinary collaboration, and public health outreach will be essential to ensure that these technologies benefit all populations equitably.

In conclusion, machine learning holds immense promise in reducing breast cancer mortality through early detection and informed decision-making. With ongoing innovation and responsible implementation, AI-driven diagnostic tools can become a cornerstone of modern oncology care.

REFERENCES

[1] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, Using Machine Learning Algorithm for Breast Cancer Risk Prediction and Diagnosis, pp. 1064–1069, 2016.
[2] D. Delen, G. Walker, and A. Kadam, Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods, Artificial Intelligence in Medicine, vol. 34, pp. 113–127, 2005.
[3] Chandrasekar R. M., Palaniammal V., and Phil M., Performance and Evaluation of Data Mining Techniques in Cancer Diagnosis, IOSR Journal of Computer Engineering (IOSR-JCE), 15(5): 39–44.
[4] Pietro Lio and Sarinder Kaur Dhillon, Predicting Factors for Survival of Breast Cancer Patients Using Machine Learning Techniques, Article no. 48, 2019.
[5] Mamta Jadhav and Zeel Thakkar, Breast Cancer Prediction Using Supervised Machine Learning Algorithms, vol. 06, 2019.
[6] Manisha Bahl and Regina Barzilay, High-Risk Breast Lesions: A Machine Learning Model to Predict Pathologic Upgrade and Reduce Unnecessary Surgical Excision, vol. 286, pp. 10–17, 2017.

[7] Cameron Wolfe, Training a Random Forest to Identify Malignant Breast Cancer Tumors, 2018.
[8] Ashish, A Random Forest Approach to Predicting Breast Cancer in Working Class Women, 2016.
[9] Z. Wang, M. Li, H. Wang, H. Jiang, and Y. Yao, Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion with CNN Deep Features, IEEE Access, vol. 7, pp. 105146–105158, 2019.
[10] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks with Small SE-ResNet Module, PLOS ONE, vol. 14, no. 3, pp. 1–21, 2019.
[11] I. A. Fondón, A. Sarmiento, A. I. García, M. Silvestre, and C. Eloy, Automatic Classification of Tissue Malignancy for Breast Carcinoma Diagnosis, Computers in Biology and Medicine, vol. 96, pp. 41–51, 2018.
[12] A. A. Saad, M. M. Kamruzzaman, M. N. Sarker, M. Alruwaili, Y. Alhwaiti, N. Alshammari, and M. H. Siddiqi, Boosting Breast Cancer Detection Using Convolutional Neural Network, Journal of Healthcare Engineering, pp. 1–7, 2021.
[13] S. S. Olofintuyi, Breast Cancer Detection with Machine Learning Approach, FUDMA Journal of Sciences (FJS), 7(2): 1–17, 2023.
[14] R. Adam, K. Dell'Aquila, L. Hodges, T. Maldijan, and T. Duong, Deep Learning Applications to Breast Cancer Detection by Magnetic Resonance Imaging: A Literature Review, Breast Cancer Research, vol. 25, pp. 1–12, 2023.
[15] J. Kiran, A. K. Muhammad, A. Majed, T. Usman, Z. Yu Dong, H. Ameer, M. Arturas, and D. Robertas, Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion, Sensors, pp. 1–23, 2022.
[16] C. Deepraj, D. Anik, D. Ajoy, S. Shreya, D. D. Ashutosh, R. M. Raghava, and M. Lakhindar, ABCanDroid: A Cloud Integrated Android App for Noninvasive Early Breast Cancer Detection Using Transfer Learning, Sensors, pp. 1–20, 2021.
[17] J. Xiao, S. Xiaolin, and Z. Xingang, Breast Cancer Identification Using Machine Learning, Mathematical Problems in Engineering, pp. 1–8, 2022.
[18] G. María, G. Miguel, R. Alicia, A. Beatriz, M. Diego, M. Amalia, F. Ana, and N. Apolonia, Genetic Variants of ANGPT1, CD39, FGF2 and MMP9 Linked to Clinical Outcome of Bevacizumab Plus Chemotherapy for Metastatic Colorectal Cancer, International Journal of Molecular Sciences, vol. 22, pp. 1–16, 2021.
[19] K. Ranpreet, H. Gholam, S. Roopak, and L. Maria, Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images, Sensors, vol. 22, pp. 1–15, 2022.
[20] S. Alvi and A. Kadam, Breast Cancer Detection Using Deep Learning and IoT Technologies, Journal of Physics Conference Series, vol. 1831(1), pp. 1–8, 2020.