IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

Mammography Using Machine Learning

Shivam Mishra, Uchit Parashar, Ashish Kumar Jha, Suhani Chauhan
Saral Garg, Pooja Chaudhary, Preeti Sharma
Raj Kumar Goel Institute of Technology, Ghaziabad, India

Abstract: *It shows evidence of breast cancer is an aspect in improvement of patient result, improves at regular interval and effective treatment, and increasing survival. Various screening methods, including imaging techniques such as regular self-examination, clinical breast examinations, and mammography, play an important role in identifying inefficiency in tissue. Mammography using Low- Power X-RAY techniques is generally considered to be the most reliable and accessible screening method, especially for women over 40 years.*

In certain cases, additional diagnostic tools such as ultrasound and MRI can be used to further examine suspected findings and improve the accuracy of the diagnosis. Sensitization, facilitating educational initiatives, and promoting routine testing are key strategies for strengthening individuals and health systems in early identification and treatment of breast cancer. Previous evidence shows that in many cases there are fewer aggressive treatment protocols and significantly improve patient predictions.

The purpose of this study focuses on using technology in machine learning to recognize breast cancer. In particular, we evaluate the effectiveness of four broad classification algorithms: SVM, RF, D T, and logistic regression. The performance of each algorithm was evaluated as the primary e valuation metric using accuracy.

After extensive experiments on data records, the support vector machine was developed as the efficient algorithm, achieving a 95% precision rate. In contrast, the random forest algorithm showed the lowest performance with 90% accuracy. These results suggest that SVM provides a high possibility of reliable and accurate recognition for breast cancer, thereby providing a promising approach to improving computer aided diagnostic systems in clinical practice..

Keywords: Breast cancer, Support Vector Machine, Logistics Regression, Random Forest, Decision Tree

I. INTRODUCTION

Breast Cancer (BC) is o common & life threatening cause affecting girls around the world. Data given by WHO, breast cancer is the more occurring & frequently diagnosed cancer in girls & leading cause of cancer related deaths. If the disease is pinched in the early stages, treatment is quite effective, so early detection and accurate diagnosis are important to improve survival. Old Diagnostics technique like breast cancer detection using ML, ultrasound, & biopsy effective, but often relate to limitations related to accuracy, cost, and accessibility. In recent years, advances in AI, particularly ML, has shown great potential to bring change the field in medical diagnosis, including breast cancer detection.

ML is a subgroup of AI that allows system to study from dataset and find decisions without making predictions. In relation to breast cancer detection, ML algorithms can find large amounts of medical dataset. This ability to process large and complex data records makes ML particularly suitable for cancer diagnosis, where accuracy and efficiency are the most important.

One of the main applications of ML in mammography is image- based diagnosed. For example, a kind of deep - flow algorithm, Folding Networks (CNNS), was often used to analyse mammographic images and recognize anomalies.

These networks can learn to distinguish benign and malignant tumours by training thousands of labelled images. In contrast to traditional methods that are strongly based on human interpretation, ML algorithms can process subtle visual

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 3, May 2025 Impact Factor: 7.67

information and provide consistent, objective evaluations. Research shows that the ML model matches the di agnostic accuracy of experienced radiologists in specific tasks.

Another promising application of machine learning is predictive modelling using structured clinical data. Monitoring technique such as SVM, decision trees, random forest, and logistics regression are used to predict breast cancer risk and to classify tumour types decided on

Characteristics like patient age, tumour size status, hormone receptors, and genetic markers. By learning from historical data, these models can provide probabilistic predictions that clinics will help with decision making and personalized treatment planning.

Unattended learning techniques such as clustering and size reduction also play a role in understanding breast cancer. These methods can reveal hidden patterns or subtypes of data records that may not be immediately recognized. For example, unattended learning can help identify new genetic signatures or phenotypic clusters associated with various predictions or responses to treatment.

Despite the promising results, the combining of ML into practice is associated with several challenges. The main issue is data quality and availability. ML models re quire large, diverse and well-approved data records to be effective generalized. However, medical data is often fragmented and unstructured, and suffers from data protection concerns. Moreover, the interpretability of the model remains an important hurdle. Clinicians need to understand how and why algorithms determine specific decisions, especially when they are at risk. Explanatory KI (XAI) research aims to solve this problem by developing models that provide prediction transparency and justification.

Ethical and legal considerations arise when providing ML systems in a healthcare system. Training data distortions can lead to differentiation of model outputs for different demographic groups. Patient equity, accountability and approval guarantees are extremely important when designing and implementing ML control diagnostic tools. A regulatory framework and strict clinical verification are necessary to build trust and ensure security.

Finally, the use of mammography is an important step towards a more efficient, more accurate, and accessible diagnostic tool. While technology is developing, interdisciplinary collaboration between data scientists, clinicians and political decision-making is essentially important to fully achieve the benefits of ML in healthcare. Further rese arch, ethical supervision, and patient centred approaches ensure that machine learning will replace human expertise in the fight against breast cancer.

II. LITERATURE REVIEW

S. No.	Author	Method	Result	Dataset	Future Scope
1.	Md. Aalim Talukder et al.[1]	XAI- DL	DenseNet169 outperforms other models, achieving remarkable accuracy scores of 99.50%, 98.80%, 97.27%, and 96.98% for BreakHis 40X, 100X, 200X, 400X, and 94.75% for the BACH dataset, with minimal false positives.	5 Dataset used BreakHis 40X, 100X, 200X, 400X, and BACH	focus on incorporating these techniques and translating the model into clinical practice for more effective breast Cancer detection.
2.	A. Rajasekhar Yadav et al.[2]	PSO-CNN Technique	99.35% accuracy, 99% sensitivity, and 98.2% specificity	7632 mammography images	Collaborate with radiologist team
3.	Vinit Kumar et al.[3]	CAMR	99.48% accuracy	DDSM dataset MIAS dataset	additional imaging
4	Omar Ramdan et.al.[4]	portable impedimetric	Detected limit 8.2 and 4.9 fg/ml	The dataset includes	The developed $CoMoO_4$@

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

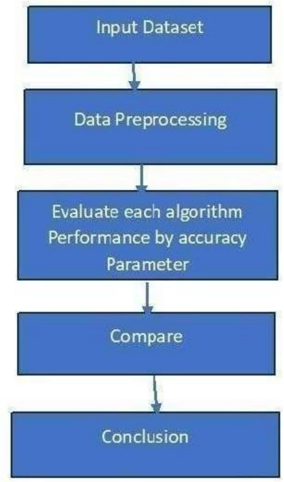
		multiplexed-immunosensing systems		detection ranges and performance metrics for breast using $CoMoO_4$.	PANI- PPy-based dual t mmunosensor holds great promise for advancing rapid, sensitive, and label-free detection of multiple breast cancer biomarkers in clinical diagnostics
5.	Cristiana MorozDubenco et al.[5]	Decision Tree algo.	accuracy of 95% & 100% precision	mini-MIAS dataset	find system's interpretability.
6.	Ahmed A.Shalaby et.al.[6]	paper-based bioluminescence ELISA	Post-implementation, SDM improved, with less decisional conflict and FCR	"Prospective dataset of 507 breast cancer survivors across 8 Dutch hospital.	optimising the implementation of the BCSPtDA by addressing barriers and facilitators
7.	Gani Esen et al.[7]	Dimesionality Reduction,Specifically Principal Component analysis	Accurate classification on data set	Wisconsin Breast Cancer Data	Advance PCA
8.	Flavio Augusto Ataliba Caldas et al.[8]	Lunit INSIGHT mammogram AI system	158 biopsies, radiologists	Calculated 617 mammograms, with biopsy no upto 104	Increase efficiency in Ai model
9.	Armin Jarahi Khameneh et al.[9]	electrochemical method	alternative to detecting breast cancer.	Data set consist of electrochemical biosensors.	Finding the circulation breast cancer.
10.	W Janni et al.[10]	custom bioinformatics classifier	Finded that 11/14 recurring patient	Samples contain 47 plasma from 38 EBC patient	Performance of assay.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, May 2025 Impact Factor: 7.67

11	Misbahu Koramar Boko Lawal et al.[11]	hybrid model (DResNet) and pretrained models	Accuracy upto 95.98% using D-ResNet-RF Validated	Dataset consist of 7783 images	Future scope includes exploring additional pretrained models, real world clinical datasets, and alternative classifiers to Enhancemodel generalizability, interpretability, clinical integration
12	Radwan Qasrawi Et al.[12]	(CLAHE)	It consist of 96% of benign	Data consist of 4103 photos	Future scope includes validating the hybrid model across multiple centers integrating it into clinical systems, and optimizing it for real-time processing and broader pathology coverage
13	Faseela Abdullakutty et al.[13]	Integration of many data	It increases robustness	multi-modal EMR dataset	Future scope includes enhancing breast cancer detection by refining multimodal fusion techniques
14	Garcia-Murillas et.al.[14]	The method used was a personalized, circulating tumor DNA	Menstivity	617 plasma samples (median8, range2-14) from 78 patients	Future scope includes integrating ultrasensitive WGS-based ctDNA assays into routine clinical practice for early detection, Personalized treatment, and monitoring of breast cancer relapse.
15	Soheil Sadr et al.[15]	use of advanced imaging technique	Gold Nano biosensors, integrated with machine learning, offer highly sensitive, specific, and non-invasive detection of breast cancer biomarkers	article does not present a specific dataset but discusses various experimental and clinical studies involving gold Nano biosensors and breast cancer biomarkers.	The future scope involves integrating gold Nano biosensors with AI and portable devices for early, precise, and accessible breast cancer diagnosis globally

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, May 2025 Impact Factor: 7.67

III. METHODOLOGY



In Fig 1 Proposed Methodology of Project

In Figure 1, the data was obtained from the Kaggle platform. Initial pre-processing involved the removal of null values, followed by comprehensive data cleaning to efficiently integrate and consistency of the Data. Four ML algorithms—SVM, Logistic Regression (LR), Random Forest, and Decision Tree—are subsequently applied create system for mammography.

Each model's performance was assessed using correct result is the primary evaluation metrics. A comparative analysis were conducted to determine the most suitable algorithm for correct and efficient breast cancer classification.

1. Support Vector Machine - SVM is a supervised ML algorithm widely used for classification work, including medical diagnosis such as breast cancer detection. It works on finding the efficient usual hyper plane i.e. efficiently divides data points of other class with max margins. algo relies on SV— critical dataset points tells define the decision limit. Support vector machine is useful in high dimensions space and works on both linear and non-linear classifications using kernel functions.

It offers strong generalization performance, especially when class separation is clear.

However, it required proper tuning & may be computationally expensive on large datasets.

2. Logistic Regression - LR is ML algo. Used primarily for binary classification work. It finds the probability that a given input data belong to a particular classes using the sigmoid function to a linear combinations of input features. The o/p is a value lies between 0 and 1, showing the likelihood of classes membership.

LR is useful, efficiently, and interpretable, makes it suitable for medical applications like breast cancer detection. It assume a linear relationship remain in range of input variables & unique outcomes. Although less effective for complex patterns, it performs well on linearly separable data.

3. Random Forest (RF)- RF is an ensemble learning algorithm used for Classification and regression work. It works by construction of occurring multiple decision tree's during training and outputs the class that the majority vote among of trees. Each tree's is built on a random subsets of the data and features, which increases efficiently and reduce over fittings. Random Forest is robust, handles both categorical and numerical data well, and is effective for datasets with missing values or noise. In breast cancer detection, it provides high accuracy and useful feature. However, shows computationally intensive and less interpretable than simpler model.

4. Decision Tree (DT) - DT is a supervised ML algo. Used for both classifications and regressions task. It work by recursive splitting the data into subsets decided on the most efficient function at each node, finding a tree-like structure. Each inside nodes tells a decision set, & every leaf node shows a result label. It easy to find result, making efficient result in medical research such as breast cancer Detection. They works on numerical and categorical dataset and need minimized data pre-processing. They are prone to over fitting, it can be mitigated using techniques like ensemble methods.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, May 2025

Impact Factor: 7.67

IV. RESULTS

1. Support Vector Machine:- From Table 2 and Figure 1, we show that the Support Vector Machine (SVM) algorithm with machine learning reaches 95% accuracy in breast cancer detection.

This phase focuses on assessing the performance of the model by

Analysing the experimental results. The reported accuracy indicates the ability of The algorithm to correctly classify cancer cases. This evaluation step is of great Importance in research to verify the effectiveness of the model and identify Areas for improvement or further testing.

2. Decision Tree - From Table 2 and Figure 1, we show that decision tree algorithms using machine learning reach 89% accuracy in breast cancer recognition. This research phase includes performance a assessments. This assessment analyses experimental results to measure the degree to which the mod el is classified. It shows 80% efficiency shows a useful level of effectiveness. This level is important for assessing the reliability of the model and for managing future improvements or comparing it with other algorithms.

3. Random Forest - Table 2 and Figure 1 show that the random algorithm achieves 96% accuracy in mammography using ML. This phase represents the working evaluation level at which the predictability of the model is evaluated by the experimental results. The 80% accuracy reflects the area of accurately find cases & shows validity of system. This step is of great importance in finding data to check the algo & decide future comparison.

4. Logistic Regression - From Table 2 and Figure 1, we show that logistic regression algorithm with machine learning reaches 94% accuracy in breast cancer recognition. This phase focuses on the level of assessment where experimental dataset is analysed to use the classification work of model. Registered 80% accuracy shows correctly predicted cases. This step is extremely important in research to verify the reliability of the model and to inform further improvements or comparisons with other methods.

Table 2 represents the algo performance for detecting breast cancer

S.No.	Algorithm	Accuracy
1.	Support Vector Machine	95%
2.	Decision Tree	89%
3.	Random Forest	96%
4.	Logistic Regression	94%

V. CONCLUSION

Breast Cancer Detection Accuracy machine reached at highest accuracy of 95%. This demonstrates the strong potential for reliable classification in medical diagnostic tasks. Opposite the RFM tell low estimated accuracy at 80%. It shows that the result was not useful for any particular datasets. The decision tree and logistics regressions algorithm showed intermediate results containing accuracy values lies in range of support vector machine & RF effective values. For future improvements, we recommend applying these models to larger and diverse data records to improve generalization. Furthermore, image- based data records such as mammograms allow for the integration of deep learning techniques, leading to accuracy and more robust diagnostic systems for detecting breast cancer.

REFERENCES

[1]. Khameneh, A. J., Rahimi, S., Abbas, M. H., Rahimi, S., Mehmandoust, S., Rastgoo, A., ... & Eskandari, V. (2024). Trends in electrochemical biosensors for the early diagnosis of breast cancer through the detection of relevant biomarkers. Chemical Physics Impact, 8, 100425.

[2]. Yadav, A. R., & Kumar, V. N. (2025). PSO-optimized fractional order CNNs for enhanced breast cancer detection. Results in Engineering, 26, 104559.

[3]. Kumar, V., Chandrashekhara, K. T., Jagini, N. P., Rajkumar, K. V., Godi, R. K., & Tumuluru, P. (2025). Enhanced breast cancer detection and classification via CAMR-Gabor filters and LSTM: A deep Learning-Based method. Egyptian Informatics Journal, 29, 100602.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, May 2025
Impact Factor: 7.67

[4]. Ramadan, O., & Hassan, R. Y. (2025). Nanostructured immunosensing system for label-free impedimetric detection of multiple breast cancer biomarkers (CEA and HER2) using CoMoO4@ PANI-PPy Nanocomposite. Biosensors and Bioelectronics: X, 23, 100596.

[5]. Moroz-Dubenco, C., Bajcsi, A., Andreica, A., & Chira, C. (2025). Towards an interpretable breast cancer detection and diagnosis system. Computers in Biology and Medicine, 185, 109520.

[6]. Shalaby, A. A., Salah, A., Ishida, A., Maeki, M., & Tokeshi, M. (2025). Paper discs in a 3D printed microplate hybrid microfluidic device for low-cost, rapid, and ultrasensitive paper- based bioluminescence detection of human epidermal growth factor receptor 2 (HER2) breast cancer biomarker. Biosensors and Bioelectronics: X, 100621.

[7]. Esen, G., Altaibek, A., Amankulov, J., Matkerim, B., & Nurtas, M. (2024). Enhancing Breast Cancer Detection with Dimensionality Reduction Techniques: A Study Using PCA and LDA on Wisconsin Breast Cancer Data. Procedia Computer Science, 251, 414-421.

[8]. Caldas, F. A. A., Caldas, H. C., Henrique, T., Jordão, P. H. F., Fernandes-Ferreira, R., Souza, D. R. S., & di Pace Bauab, S. (2025). Evaluating the performance of artificial intelligence and radiologists accuracy in breast cancer detection in screening mammography across breast densities. European Journal of Radiology Artificial intelligence,2,100013.

[9]. Trends in electrochemical biosensors for the early diagnosis of breast cancer through the detection of relevant biomarkers

[10]. Detection of minimal residual disease and prediction of recurrence in breast cancer using a plasma-only circulating tumor DNA assay

[11]. Lawal, M. K. B., Almousa, M., Ibrahim, A. U., Pwavodi, P. C., Usman, A. G., & Aloraini, B. (2025). Artificial intelligent-powered detection of breast cancer. Journal of Radiation Research and Applied Sciences, 18(2), 101422.

[12]. Qasrawi, R., Daraghmeh, O., Thwib, S., Qdaih, I., Issa, G., Polo, S. V., ... & Atari, S. (2025). Advancing Breast Cancer Detection in Ultrasound Images Using a Novel Hybrid Ensemble Deep Learning Model. Intelligence-Based Medicine, 100222.

[13]. Abdullakutty, F., Akbari, Y., Al-Maadeed, S., Bouridane, A., Talaat, I. M., & Hamoudi, R. (2024). Towards improved breast cancer detection via multi-modal fusion and dimensionality adjustment. Computational and Structural Biotechnology Reports, 1, 100019.

[14]. Garcia-Murillas, I., Abbott, C. W., Cutts, R. J., Boyle, S. M., Pugh, J., Keough, K. C., ... & Turner, N. C. (2025). Whole genome sequencing powered ctDNA sequencing for breast cancer detection. Annals of Oncology.

[15]. Amethiya, Y., Pipariya, P., Patel, S., & Shah, M. (2022). Comparative analysis of breast cancer detection using machine learning and biosensors. Intelligent Medicine, 2(2), 69-81.

[16]. Marchi, R., Hau, S., Suryaningrum, K. M., & Yunanda, R. (2024). Comparing YOLOv8 and YOLOv9 Algorithm on Breast Cancer Detection Case. Procedia Computer Science, 245, 239246.

[17]. Smilkou, S., Ntzifa, A., Stergiopoulou, D., Georgoulias, V., & Lianidou, E. (2024). Multiplex detection of ten ESR1 mutations and AKT1 E17K in breast cancer using digital PCR. The Journal of Liquid Biopsy, 5, 100154.