**AI in Breast Cancer Screening: A Comparative Analysis of Current Approach (ML)**

K R Shwetha1, R Shilpa2, Rajesh I S3, Sowmya V L4

[1shwethakr1704@gmail.com](mailto:1shwethakr1704@gmail.com), [2shilparamineni21@gmail.com](mailto:2shilparamineni21@gmail.com), [3rajeshaiml@bmsit.in](mailto:3rajeshaiml@bmsit.in), [4soumya8.vl@gmail.com](mailto:4soumya8.vl@gmail.com)

1,2,3,4 AIML Department, BMS Institute of Technology and Management, Bengaluru

**Abstract.** Breast cancer is one of the primary causes of death among women in developing countries, highlighting the urgent need for effective methods of early diagnosis and treatment. This disease, which originates in breast tissue, is generally categorized into two main types: Invasive Ductal Carcinoma (IDC) and Ductal Carcinoma in Situ (DCIS). In recent years, advancements in Artificial Intelligence (AI) and Machine Learning (ML) have significantly transformed approaches to diagnosis and prevention, particularly with the use of technologies like Magnetic Resonance Imaging (MRI) and Convolutional Neural Networks (CNNs). This paper examines the various stages, classifications, and symptoms of breast cancer, emphasizing the importance of mammography in detection. It also provides an in-depth comparison of AI-based methods used in breast cancer research, analyzing their advantages, challenges, and the datasets employed. The insights offered aim to guide future developments in early detection techniques and predictive modeling for this critical health issue.

**Keywords:** Breast Cancer, mammography, Artificial Intelligence, Machine Learning, MRI, CNN.

1. Introduction

Cancer is the term used to describe the abnormal development and dissemination of cells inside the body. Cancer can spread to far-off areas of the body and endanger life if treatment is not received. It specifically begins in the cells of the breast tissue. Female gender, advancing age, genetic susceptibility, and lifestyle factors including food and physical exercise are major risk factors for breast cancer [6]. Early detection and routine screening are key to better outcomes. In the twenty-first century, breast cancer is a major health issue for women. It continues to rank among the most often discovered tumors and significantly affects mental and physical health [9]. Women are diagnosed with breast cancer in about 99 percent of instances. males still receive between 0.5 and 1% of all breast cancer diagnoses, despite the fact that breast cancer in males is less frequent. In 2022, breast cancer killed 670,000 individuals worldwide and affected around 2.3 million women, making it a serious global health crisis. In situ refers to the early stage of the disease when the cancerous cells are restricted to their original site and not yet invasive. The cancer may become invasive as it advances, infiltrating the surrounding breast tissue. Tumors may develop as a result of this invasion and show up in the breast as lumps or dense spots, as seen in Fig. 1. Currently, mammography, ultrasound, X-rays, and other methods are employed for breast cancer screening. Artificial Intelligence (AI) developments have enhanced preventative and diagnostic strategies. [7]

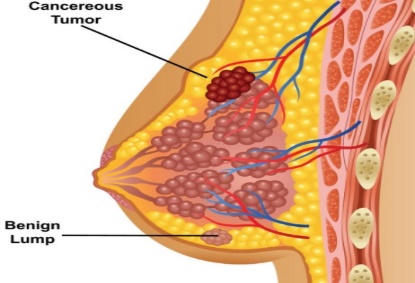


Fig. 1. Attack of breast cancer

A prospective answer to technological developments that could improve the effectiveness and precision of mammography interpretation is artificial intelligence (AI). Large datasets of mammogram pictures and the diagnoses that go with them may be applied to teach AI systems to recognize patterns linked to breast cancer. Then, by highlighting questionable regions for more examination, these algorithms might help radiologists prioritize cases for prompt attention and possibly lower the possibility of human error. Two popular methods are machine learning and deep learning for breast cancer detection. Mammograms are the main screening method, and machine learning aids in diagnosis. Since symptoms frequently appear later, routine screening is crucial. Al uses pattern recognition and data processing to provide precise detection.

* 1. Classification of Breast Cancer

Invasive breast cancer and non-invasive breast cancer are the two main classifications for breast cancer. Non-invasive breast cancer is found in the milk glands; the surrounding breast tissues are unaffected. Two non-invasive types of breast cancer exist, such as:

* **Ductal Carcinoma in Situ (DCIS):** Malignant cells that are limited to the breast ducts are typically detected by mammograms.
* **Lobular Carcinoma in Situ (LCIS):** During biopsy operations performed for various purposes, malignant cells found in the breast lobules are often found by accident. [8]

The invasive types of breast cancer are as follows:

* **Invasive Lobular Carcinoma (ILC):** Under a microscope, this cancer may be found on both sides after infiltrating the surrounding breast tissues.   
  One uncommon and aggressive type of breast cancer that resembles a rash or inflamed breast tissue is called inflammatory breast cancer (IBC). It blocks the lymphatic veins in the skin, and the diagnosis is typically made under a microscope.
* **Paget's Disease:** This rare type of breast cancer can start on the nipple and progress to the areola. A biopsy and microscopic analysis are frequently carried out to diagnose it.
* **Invasive Ductal Carcinoma (IDC):** If the tumor is big or visible on a mammography, cancer cells may be felt when they invade the surrounding breast tissues. IDC is responsible for 81% of all cases of breast cancer.
  1. General Machine Learning Algorithms to Detect Breast Cancer

The objective is to identify the optimal algorithm for breast cancer detection. This can be accomplished by applying classifiers for machine learning to Wisconsin breast cancer diagnosis dataset. After that, the outcomes are assessed to ascertain which model produces the best accuracy [10]. Below is a description of the many machine learning algorithms:

* **Support Vector Machine (SVM):** This classifier separates datasets into groups by determining the maximum marginal hyperplane (MMH) using the nearest data points. It is efficient in high-dimensional spaces.
* **Random Forests:** A method of classification that builds multiple decision trees and predicts based on the majority vote is an ensemble technique.
* **K-Nearest Neighbours (K-NN):** It is a technique of supervised classification for labeling new points using the nearest neighbors.
* **Logistic Regression:** One of the most common modeling techniques to predict binary or binary classes dependent variable in medical diagnostics.
* **Decision Tree C4.5:** A decision tree is appropriate for classification problems, such as breast cancer diagnosis, where different conditions will recursively divide the dataset to construct the tree.
* **Artificial Neural Networks (ANNs):** Being made of nodes that can communicate with each other through weights, ANNs discover the latent pattern of data and assist in detecting breast cancer.

1. Dataset Acquisition

Mammography, a specialist imaging method that views breast tissue using low-dose X-rays, is frequently used to diagnose breast cancer. Other imaging modalities such as ultrasound and MRI can be used for further evaluation or when mammograms are inconclusive. Several online databases offer mammographic images for research and educational purposes, as listed in Table 1.

Table 1. Usable mammographic image repositories

|  |  |
| --- | --- |
| **Datasets** | **Link to the archive download** |
| DDSM | http://www.eng.usf.edu/cvprg/mammography/database.html |
| CBIS-DDSM | https://www.kaggle.com/datasets/awsaf49/cbis-ddsm-breast-cancer-image-dataset |
| BCD | https://www.kaggle.com/datasets/anwarsalem/dataset-bcd-mammography-images-out |

1. Literature Survey

Through sophisticated methods like machine learning, which evaluate medical images to accurately identify tumors, artificial intelligence is used in the identification of breast cancer. These techniques help radiologists by offering accurate and timely interpretations of mammography and other imaging modalities, improving screening procedures, and promoting early diagnosis. The numerous methods, techniques, and datasets that various writers have used to identify breast cancer are detailed in Table 2. We also highlight the many advantages and challenges associated with using these methods. The literature survey's description and illustration are provided in Table 3.

Table 2. Literature survey on breast cancer detection

|  |  |  |  |
| --- | --- | --- | --- |
| **Authors** | **Methods/ Techniques** | **Dataset** | **Remarks** |
| Kwak, D. et al. (2023) | - Image classification models.  - Image segmentation models.  - Loss functions.  - Data augmentation. | - ACECR  - MGH Patient Mammograms.  - INbreast. | - It explores using deep learning methods to improve breast cancer treatment diagnosis through image recognition.  - It demonstrates how advanced algorithms can enhance the accuracy of detecting cancerous lesions in medical images. |
| Nneji, G. U. et al. (2023) | - Convolutional Neural Networks (CNNs).  - Histopathological image analysis.  - Deep learning and lightweight architectures | - BreaKHis Dataset. | * It presents an efficient lightweight separable convolution network for identifying breast cancer histopathology. * It offers a solution that balances computational efficiency and diagnostic accuracy. |
| Sechopoulos, I. et al. (2021) | - Deep learning-based detection.  - Feature extraction via CNNs.  - Digital tomosynthesis enhancement. | - INbreast, CBIS-DDSM. | - It stresses the truly important role of AI in reducing false positives substantially, improving early detection greatly, and assisting radiologists deeply in diagnosis.  - It identifies dataset biases as some key challenges. It also identifies the need for complete AI model interpretability and many large-scale validation studies. |
| Pacilè, S. et al. (2020) | - Image preprocessing and enhancement.  - AI-driven feature extraction.  - Multi-model diagnostic integration. | - Digital Database for Screening Mammography (DDSM) | - Points out AI’s meaningful role in comprehensively reducing the workload for all radiologists, as it maintains or further improves diagnostic precision.  - It intently focuses on AI's simultaneous use in clinical settings, consequently remaining relevant for mammography's real implementation. |
| Gardezi, S. J. S. et al. (2019) | - Machine Learning (ML).  - Deep Learning (DL) techniques. | - Multiview Mammogram Datasets. | - Highlights machine learning advancements, focusing on sensitivity, specificity, and dataset diversity.  - Identifies challenges such as imaging variability and the need for interpretability. |
| Nakahori, R. et al. (2015) | - Imaging techniques.  - Biopsy and diagnosis.  - Surgical intervention. | - The Cancer Imaging Archive (TCIA).  - Clinical Proteomic Tumor Analysis Consortium (CPTAC).  - Gene Expression Omnibus (GEO)- | * It emphasizes the importance of recognizing the potential for malignancy in silicone granulomas, which may be overlooked in routine assessments.   - It needs more investigation into the long-term consequences of silicone on breast tissue and the need for careful monitoring in patients with silicone implants. |

An analysis of literature reveals that there has been a major shift in detection methods for breast cancer, from olden conventional procedures to very recent state-of-the-art techniques. An example of the recent studies was application of deep learning for improved accuracy in detection. Sechopoulos et al. (2021) and Pacilè et al. (2020) used feature extraction and diagnostic integration to assist a radiologist under false-positive reduction. In 2023, lightweight neural networks were used by Nneji et al. in interpreting histopathology images while managing efficiency in processing and precision in diagnosis. These contributions clearly show how technology is rapidly advancing toward improving detection of breast cancer while addressing other critical issues, such as dataset variability and model efficiency, both of which are among the leading areas of research.

Table 3. Illustration of Literature Survey

|  |  |
| --- | --- |
| **Description** | **Results** |
| Kwak, D. et al. (2023). The graphic displays examples from three medical image datasets and reveals a significant disparity in class between the tumor and non-tumor regions. The background-to-foreground ratio is 85.5:14.5 in the ultrasound dataset and 99.56:0.44 in the X-ray dataset. The problem of class imbalance has a detrimental effect on segmentation performance (Fig. 2). [5] | Fig. 2. (a) X-Ray (b) Ultrasound (c) Histopathology |
| Nneji, G. U. et al. (2023). At different magnifications, the illustration shows both benign and cancerous BreakHis pictures. Over 2400 benign and 5400 malignant breast cancer histopathology pictures from over 80 individuals, taken at 40×, 100×, 200×, and 400× magnifications, are included in the BreakHis dataset (Fig. 3). [3] | Fig. 3. (a) Benign tumor (b) Malignant tumor |
| Sechopoulos, I. et al. (2021). A 44-year-old woman's right breast has invasive ductal carcinoma, according to her digital mammography. A superimposed heat map in the accompanying graphic shows the precise region that had the biggest influence on the final classification choice (Fig. 4). [11] | Fig. 4. Mammography shows invasive ductal carcinoma in a 44-year-old, with a heat map marking the key diagnostic area. |
| Pacilè, S. et al. (2020).  Mammograms were performed on a 51-year-old woman diagnosed with invasive lobular carcinoma. Mediomedial oblique and cranio-caudal views (upper panels) show focal asymmetry and distortion in the left breast. Close-up view confirms a mass measuring 1.5 cm in the upper outer quadrant on the craniocaudal view (circled) (Fig. 5). [12] | Fig. 5. A 51-year-old woman's initial mammogram missed invasive ductal carcinoma. A year later, follow-up revealed a 1.5-cm cancerous mass in the left breast. |
| Gardezi, S. J. S. et al. (2019). Pectoral Muscle (PM) identification in the MLO images is crucial in order to avoid false-positive findings and obstruct any lesion diagnosis (Fig. 6) [1]. So, pre-processing of mammograms in CAD systems would assist in clearance from any annotations, labels, or noise, thereby facilitating abnormality detection. | Fig. 6. (a) 1024 x 1024 original mammography image. (b) Annotation removal through pre-processing. (c) Removal of Pectoral Muscle (PM) by expanding regions. (c) Adaptive segmentation for PM removal |
| Nakahori, R. et al. (2015). (a) A low Signal Intensity (SI) mass (3.5 × 3.2 × 4.0 cm) is seen on a T1-weighted MRI, retracting the right nipple. Adipose tissue exhibits low SI oval forms, and the pectoralis major muscle is disrupted (arrows). (b) A T2-weighted MRI scan displays a mixed high and low SI mass. Central low intensity and varied early enhancement are visible in MRI with dynamic contrast (c, d) (c: early phase, d: delayed phase) (Fig 7). [4] | Fig. 7. (a) Initial T1-Weighted MRI (b) T2 Weighted MRI (c) Early Contrast-Enhanced MRI (d) Delayed Contrast-Enhanced MRI |

1. Issues in Detection of Breast Cancer

In medical diagnostics, breast cancer detection is a crucial field, but it presents a number of difficulties that may compromise the precision and dependability of screening techniques like mammography. These difficulties are caused by the intricacies of deciphering mammograms as well as the limits of the technology itself. Some of the main challenges in detecting breast cancer are listed below:

* **Dense Breast Tissue:** Because malignant tumors can be hidden by dense tissue, making identification challenging, women with dense breast tissue frequently have a higher risk of false negative results.
* **High False Positives and False Negatives:** The accuracy of diagnosis may be impacted by mammograms' tendency to create false positives, which indicate non-cancerous abnormalities, and false negatives, which miss genuine cancers.
* **Variability in Radiologist Interpretation:** Different radiologists may interpret mammography in quite different ways, which could result in inconsistent diagnoses.
* **Small and Early-Stage Tumors:** Because small or early-stage tumors are often overlooked or mimic benign illnesses, they can be particularly difficult to detect.
* **Calcifications and Mass Detection:** It can be challenging to differentiate between normal and cancerous masses or calcifications, frequently necessitating additional testing.
* **Limited Generalization of AI Models:** Although AI models show promise, variations in training data may make it difficult for them to generalize across various demographics, imaging systems, or organizations.
* **Screening Limitations in Younger Women:** Because younger women have denser breast tissue, which makes early diagnosis more challenging, mammograms are less successful in identifying cancer in these women.

1. Screening Methods to Detect Breast Cancer

* Mammography, first developed by Salomon in 1913, uses X-rays to produce two-dimensional images of the breast in order to detect cancer, as seen in Fig. 8. Masses, microcalcifications, architectural deformities, and asymmetries are all radiographic indicators of breast cancer. The quality of mammograms has significantly improved due to technological developments. Ductal Carcinoma in Situ (DCIS) and other small lesions can be detected by mammography.

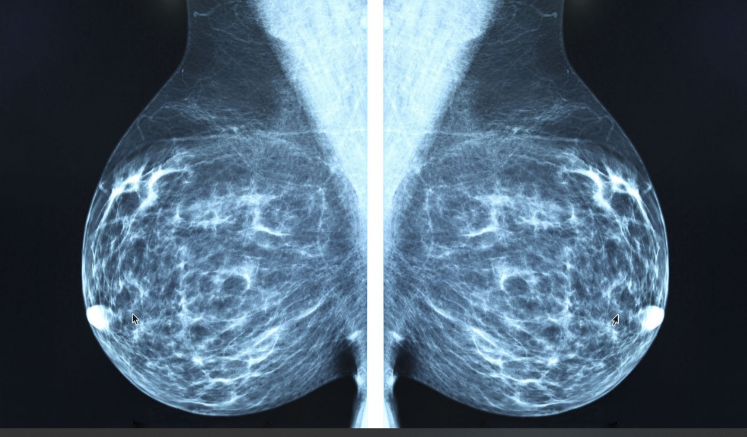


Fig. 8. Screening Using X-Ray

* Non-mammographic imaging techniques can be applied to screen for breast cancer either in addition to or instead of mammography. These techniques offer opposing perspectives and can be particularly useful in some circumstances, such as women with dense breast tissue or ambiguous mammograms. Important non-mammography imaging techniques include DBT, magnetic resonance imaging (MRI), ultrasound, as illustrated in Fig. 9.



Fig. 9. Screening Using Ultrasound

1. A Broad Method for Identifying Breast Cancer

The generalized approach to detect breast cancer includes:

* **Prediction Phase**: Model performance is assessed by comparing its predicted outcomes to actual data values on a separate test dataset. The correctness and practicality of the model in real-world situations are assessed using this evaluation. Fig. 10 displays the flow chart that outlines the overall process for detecting breast cancer.
* **Data preparation:** Data preparation involves loading the dataset on breast cancer, standardizing or normalizing characteristics, and separating features (X) and labels (Y).
* **Model Training and Evaluation:** Makes use of a variety of machine learning classifiers, including LR, SVC, DT, KNN, and RF. After training on a training set, models are tested on a test set. Algorithms such as DT and RF undergo feature significance analysis.
* **Feature selection:** Feature selection involves selecting useful and discriminative features using techniques including low-variance feature removal, univariate feature selection, and recursive feature elimination.
* **Model Evaluation:** A model's performance is frequently evaluated using measures like F1-score, recall, accuracy, and precision.

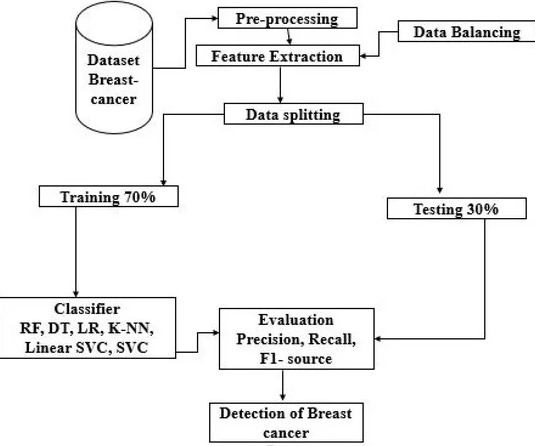


Fig. 10. Generalised Flow Chart

1. Conclusion

This paper examines the symptoms, stages, and early detection of breast cancer while also providing insight into the potential benefits of AI and ML in diagnosis and prevention. Because ML algorithms can distinguish between benign and malignant cancers, fewer needless procedures will be required. Through pattern recognition and data analysis often using mammogram images, AI enables precise detection. This paper examines the several datasets and methodologies used by academics, as well as the benefits and drawbacks of each. By using two stages, training with existing data and testing with fresh data, ML models are built to find patterns in data. But there are still issues with handling imbalanced datasets, enhancing model interpretability, and testing models in actual clinical situations. The accuracy and durability of diagnosis can be improved by integrating several data types, such as imaging, genomics, and clinical data. Patient outcomes can be greatly enhanced by putting an emphasis on early identification and precision treatment through the use of sensitive and targeted models. Another exciting area of research is creating models that use molecular profiles to predict therapeutic response and tailor treatment. Machine learning technologies for the detection and management of breast cancer can advance and enhance patient outcomes by tackling these problems.

1. Acknowledgement

We would like to express our gratitude to the experts and scholars whose work served as the basis for our review. They have made significant contributions to the profession by creating and improving breast cancer detection techniques. We also thank the providers of datasets that have been crucial to the advancement of this field of study, including ACECR, INbreast, DDSM, and others. We also thank our peers and mentors for their support and advice, whose insightful criticism has helped us improve this work.

1. References

[1] Gardezi, S. J. S. et al. (2019). Breast cancer detection and diagnosis using mammographic data: Systematic review. Journal of Medical Internet Research 21(7), e14464. <https://doi.org/10.2196/14464>

[2] Tahmoush, D. (2009). Image similarity to improve the classification of breast cancer images. Algorithms 2(4), 1503–1525. <https://doi.org/10.3390/a2041503>

[3] Nneji, G. U. et al. (2023). Lightweight separable convolution network for breast cancer histopathological identification. Diagnostics 13(2), 299. <https://doi.org/10.3390/diagnostics13020299>

[4] Nakahori, R. et al. (2015). Breast carcinoma originating from a silicone granuloma: A case report. World Journal of Surgical Oncology 13, 72. <https://doi.org/10.1186/s12957-015-0509-6>

[5] Kwak, D. et al. (2023). Rethinking breast cancer diagnosis through deep learning-based image recognition. Sensors 23(4), 2307. <https://doi.org/10.3390/s23042307>

[6] Dileep, G., & Gianchandani Gyani, S. G. (2022). Artificial intelligence in breast cancer screening and diagnosis. Cureus 14(10), e30318. <https://doi.org/10.7759/cureus.30318>

**[7] Gupta, R., & Bilal, M. (2023).** Breast cancer screening using artificial intelligence techniques: Enhancing biochemical insights and diagnostic accuracy. In Chemistry Proceedings, 14(1), 41. <https://doi.org/10.3390/ecsoc-27-16121>

[8] Khalid, A. et al. (2023). Breast cancer detection and prevention using machine learning. Diagnostics (Basel, Switzerland) 13(19), 3113. <https://doi.org/10.3390/diagnostics13193113>

[9] Islam, M. M. et al. (2020). Breast cancer prediction: A comparative study using machine learning techniques. SN Computer Science 1(5). <https://doi.org/10.1007/s42979-020-00305-w>

[10] Harinishree, M. S. et al. (2021). Detection of breast cancer using machine learning algorithms – A survey. Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2021. <https://doi.org/10.1109/ICCMC51019.2021.9418488>

[11] Sechopoulos, I. et al. (2021). Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: State of the art. Seminars in Cancer Biology 72, 214–225. <https://doi.org/10.1016/j.semcancer.2020.06.002>

[12] Pacilè, S. et al. (2020). Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool. Radiology: Artificial Intelligence 2(6), e190208. <https://doi.org/10.1148/ryai.2020190208>