AI-Powered Breast Cancer Detection: Exploring Deep Learning and Machine Learning for Early Diagnosis

Manoswini Sahoo,1, a) Priyadarshini Sahoo,2, b) Bijay Kumar Paikaray,3, c) Chandrakant Mallick,4, d) V.S. Damodharan,5, e)

1,2,3Centre for Data Science, Department of Computer Science and Engineering, Siksha 'O' Anusandhan (Deemed to be) University, Odisha, India

4Department of Computer Science & Engineering, GITA Autonomous College, Bhubaneswar, India

5Department Chair Business, Abu Dhabi Vocational Education & Training Institute, Abu Dhabi, United Arab Emirates,

c) Corresponding author: bijaykumarpaikaray@soa.ac.in

a)manoswinisahoo14@gmail.com

b)priyadarshinirimi4@gmail.com

d)ckmallick@gmail.com

e)sriramdams@gmail.com

**Abstract:** Breast cancer is one of the most common and dangerous types of cancer in the world. It affects a lot of women, but it can also affect men. Timely recognition and identification are crucial for reducing mortality rates. Artificial Intelligence (AI) and Machine Learning (ML) methods are becoming more used in medical imaging to help doctors find and classify breast cancer. Deep Learning (DL) methods have been quite successful at finding problems in mammograms, ultrasounds, and MRI scans. This chapter looks closely at different ways to use machine learning and deep learning to find breast cancer, focusing on their pros and cons. Algorithmic prejudice is a big problem in AI-driven healthcare. It happens a lot when datasets aren't balanced, certain groups aren't well represented, or technology isn't up to par. Unaddressed prejudices might lead to misdiagnosis or unequal healthcare outcomes. This paper evaluates the current advancements in AI-driven breast cancer detection, addresses sources of AI bias, and delineates ways for their mitigation to ensure equitable and dependable diagnostic results.

**Keywords:** Deep learning, Medical imaging, Breast cancer, Machine learning, AI bias

**INTRODUCTION**

In the recent world, breast cancer (BC) is the most common and deadliest cancer widely spread globally. A malignant tumor that develops in the breast’s glandular epithelium is called BC, Yao Lu at. [1]. According to the IARC, Breast cancer is one of the most common cancers in women and its possible factors involve genetic, age, lifestyle choices and environmental factors. In 2020, the IARC documented about 2.3 million new cases of BC globally, establishing it as the most frequently identified form of cancer. The worldwide fatality rate for BC stood at approximately 685,000 deaths during the same year. Lesions play a critical role in the context of BC as they are essential for identifying and understanding the disease. These lesions display abnormal changes in tissue in different ways, aiding in the distinction between different types of BC. Common risk factors for BC women after the age of 55 especially risk increasing, Risk may rise if close relatives have a BC history, Gene mutations such as those in BRCA1 or BRCA2 considerably increase risk.

**Type of Breast Cancer**

Several types of breast cancer occur, but among them the most recurrent type is Invasive Ductal Carcinoma (IDC), which develops in the milk ducts of the breast and can spread to nearby tissues. Approximately 80% of incidences of breast cancer are related to IDC. The second variant, Invasive Lobular Carcinoma (ILC), originates in the lobules, the milk-producing glands of the breast during lactation. Another type is Ductal Carcinoma in Situ (DCIS), a non-invasive carcinoma that originates in the milk ducts but remains confined without spreading to the surrounding tissue.

**Breast Cancer Images**

Various modalities of BC imaging are employed, including X-ray mammography, 3D ultrasound, computer-aided detection, computer-aided detection, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET) scans. The majority of radiologists continue to require support from computer-aided diagnosis systems for the detection and categorization of BC. Over the years, the identification of breast cancer has improved a lot, thanks to a number of methods that help doctors find it earlier and help patients get better. Mammography, a conventional technique, is the predominant screening modality for the identification of early-stage cancers. However, its limitations in sensitivity, especially for dense breast tissue, have led to the creation of additional methods. Ultrasound and MRI are being used more and more to help see things better, especially in people who are at higher risk. Recent advancements in digital tomosynthesis, which generates three-dimensional pictures of the breast, have shown promise in improving diagnostic precision. Researchers are also looking into molecular imaging techniques like Positron Emission Tomography (PET) to see whether they can find more aggressive types of cancer. AI and ML techniques are also becoming more common, which makes it possible to automate image processing and improve diagnosis accuracy. All of these improvements aim to make early detection better, reduce false positives, and tailor screening procedures to each person's risk profile.

In the age of AI, DL, and ML, these methods are being used more and more in medical imaging, and they are getting very good results. Unlike traditional diagnostic methods that depend solely on visual assessment and the expertise of radiologists, AI-driven systems can process vast amounts of imaging data and uncover intricate patterns that human observers might overlook. This skill greatly improves the accuracy of finding cancerous cells in the body and makes it easier to find breast cancer (BC) early. Early identification is vital, since it directly influences treatment options and patient survival rates. Deep learning and machine learning algorithms have shown that they can automatically sort out anomalies in mammograms, accurately telling the difference between benign and malignant tumors. Deep learning approaches, especially convolutional neural networks (CNNs), stand out among all AI methods because they can solve problems quickly, extract features in a hierarchical way, and combine data from many sources, such as mammography, ultrasound, and MRI pictures.

Moreover, deep learning models persistently develop and adjust as additional data is acquired, hence enhancing their reliability over time. For healthcare professionals, these improvements indicate that AI complements their expertise rather than supplants it, acting as a secondary reviewer, decreasing diagnostic inaccuracies, limiting false positives and negatives, and aiding in intricate decision-making. The incorporation of AI in breast cancer detection enables physicians to augment diagnostic precision, provide customized care approaches, and improve overall patient outcomes.

**Contributions of the study**

The main contributions of this chapter are:

* + Introduce a comprehensive overview of ML/DL methods in breast cancer detection.
  + Highlight role of CNN, RNN, KNN, and Random Forest in medical imaging.
  + Provide comparative analysis of datasets and algorithms.
  + Discuss AI bias implications in breast cancer detection.
  + Suggest strategies for reducing bias in AI-enabled medical systems.

The rest of the chapter is organized as follows: Section 2 reviews related literature. Section 3 discusses datasets. Section 4 presents methods and materials. Section 5 highlights prognostic tools. Section 6 provides testing and discussion. Section 7 introduces AI-bias issues. Finally, Section 8 concludes the chapter with key findings and future research directions.

**LITERATURE REVIEW**

S. Ranjana et al. [13] found several techniques to identify and categorize BC, also included a review of DL. It examines DL, image processing, ML, and datamining techniques. At the end review study finds that it outperforms other methods. Zhou X et al. [14] compiled various methods for classifying BC through HIV, utilizing multiple ANN architectures. The authors categorized their work according to the dataset employed. The items were arranged in ascending chronological order. Pang T et al. [15] investigated recent studies employing various imaging modalities and DL techniques for BC detection. The grouping was based on the dataset, construction, implementation, and attributes. The study focused on DL architectures tailored for three types of breast imaging: MRI, mammography, and ultrasound. In their work, they decided to provide the most recent information on BC imaging utilizing DLR-based computer-aided detection (CAD) systems. The researchers employed a convolutional neural network for classification on a proprietary dataset. Ali Bou Nassif et al. [16] divided the study into two groups based on multiclass and binary classification. The findings indicate that the binary classification technique achieved a superior performance compared to the multiclass classification, boasting an accuracy rating of 99.7%. Wisesty UN et al. [17] focused on employing gene mutations for the diagnosis of BC. To ascertain the activity of a malignancy, they clarified that the gene prediction classification phase endeavours to perform gene annotation, gene discovery, and gene mutation detection. They concluded that a range of methodologies, including SVMs, regression, and DL methods, were utilized. Tahmooresi [18] This paper proposes a hybrid ML model that integrates Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Nearest Neighbours (KNN), and Decision Tree (DT) methodologies. The results indicate that, among all the models, SVM had the highest accuracy. The SVM achieved a maximum accuracy of 99.8%, which might potentially be elevated to 100%. Mahmood M et al. [19] demonstrated that neural networks are effective in diagnosing conditions, especially in their initial phases. Their research reveals that the majority of neural networks have shown potential in identifying cancerous cells. However, to prepare the image. Xiaomin Zhou and colleagues [20][21][22] Artificial Neural Networks are crucial for the detection and management of BC. Artificial Neural Network techniques are commonly employed in the segmentation and categorization of breast histopathology pictures to enhance the accuracy of Breast Histopathology Image Analysis. Some other studies in the literature have been summarised in table 1.

**TABLE 1**: Summary of some related studies in the literature

| **Sl. No.** | **Author/ Reference** | **Year** | **Datasets** | **Techniques** | **Performance** |
| --- | --- | --- | --- | --- | --- |
| 1. | Zhao, et al. [6] | 2021 | Local training data | SMORE  (based on CNNs) | K = 0.73 |
| 2. | Saxena, et al. [7] | 2020 | BreaKHis | CNN | Acc= 892.61% |
| 3. | Liang, et al. [8] | 2020 | Local data | CNN | DSC= 0.73 |
| 4. | Mo et al. [9] | 2023 | BUSI | Vit | AUC=0.898 |
| 5. | Saber, et al. [10] | 2021 | Public | Resnet50  VGG16 | Acc = 96%, and 94% |
| 6. | Nangalia et al. [11] | 2022 | Public | KNN, SVM | Acc = 78% |
| 7. | Mallick et al. [12] | 2023 | WDBC | Random Forest | Acc= 98.63% |
| 8. | Swain et al. [13] | 2020 | Private | SVM | Acc= 98.13% |
| 9. | Priya et al. [14] | 2024 | ImageNet | EfficientNetB0 | Acc= 94.49% |
| 10. | Umer et al. [15] | 2023 | Custom | Feature Selection Deep Learning | Acc= 92.7% |
| 11 | Naji et al. [16] | 2021 | WDBC | Support Vector Machine | Acc= 97.2% |
| 12. | Nemade et al. [17] | 2023 | WDBC | Decision Tree and XGBoost | Acc= 97%  AUC=0.99  (XGBoost) |
| 13. | Rabiei et al. [18] | 2022 | Motamed cancer institute, Iran | Random Forest | Acc=80%, AUC= 0.56 |
| 14. | Mallick et al. [19] | 2025 | WDBC | Random Forest + PCA | Acc = 98.24% |
| 15. | Qian et al. [20] | 2025 | Mammography,Ultrasound | Multimodal Model | Acc = 92.7% |
| 16. | Singh et al. [21] | 2024 | WDBC | Optimized ML-driven CAD | Acc = 97.96% |
| 17. | Dash et al. [22] | 2022 | Benign Breast Tumor Dataset | VGG-16, MobileNetV2, | Acc= 97.22%, 98.61 |

The reviewed literature demonstrates the rapid progress of AI-based approaches in breast cancer detection. Ranjana et al. [13] and Zhou et al. [14] emphasized that machine learning and artificial neural networks are effective in identifying and categorizing breast lesions, while Pang et al. [15] highlighted the utility of deep learning radiomics across MRI, mammography, and ultrasound imaging. Mo et al. [9] employed a Vision Transformer (ViT) on the BUSI dataset and reported strong AUC performance, whereas Saber et al. [10] applied transfer learning with ResNet50 and VGG models to achieve accuracies above 94%. Other studies such as Nassif et al. [16] and Wisesty et al. [17] explored binary vs. multiclass classification and gene mutation-based diagnosis, respectively, showing that binary models tend to perform more reliably. More recent works, including Mallick et al. [12, 19], Nemade et al. [17], and Qian et al. [20], proposed ensemble and multimodal frameworks that integrate clinical, genetic, and imaging features, achieving accuracy levels above 97%. Despite these advancements, several studies note challenges such as dataset imbalance, demographic underrepresentation, and limited generalization across populations, raising concerns of algorithmic bias. Collectively, the surveyed works establish CNNs, hybrid ML models, and transformer-based methods as highly promising for early breast cancer detection, while underscoring the need for diverse datasets and fairness-aware strategies to ensure equitable healthcare outcomes [23].

**DATASET**

This dataset delivers a plethora of knowledge for developing and assessing DL techniques for BC diagnosis. The collection's versatility for algorithm development and educational initiatives is enhanced by the presence of augmented X-rays.

**Commonly Used Mammography Datasets**

A variety of imaging techniques, such as histology imaging, mammography, and ultrasound imaging. Integrating the detection result from all three imaging techniques is the most effective approach to enhance BC diagnosis. This section presents several commonly used datasets for evaluating the accuracy of breast cancer detection across various imaging techniques.

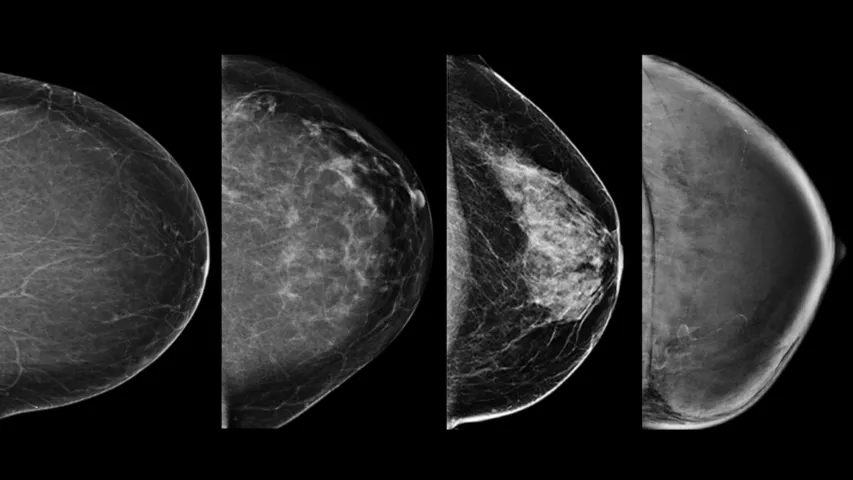
Mammography is a widely utilized imaging technique for the early detection of breast cancer. Mammography images can detect various lesions, including masses and micro calcifications

* + 1. **Mini-MIAS**: One widely recognized set of digital mammography visuals used in BC research is the mini-MIAS dataset. A total number of 322 pictures in this dataset each bring a dimension of 200 microns with a pixel size of 1024×1024. The picture shows the breast through MLO perspectives, and every image is supported by information about the breast tissue, where it is and the identification of any abnormalities. The mini-MIAS dataset is frequently utilized to evaluate how accurately ml algorithms and CAD systems detection BC.
    2. **Digital Database for Screening Mammography**: Another one popular dataset of digital mammography visuals used in BC research is DDSM. Here two pictures of the left and right breasts are present in each of the 2620 instances in a dataset, which come from 1877 patientsEach image has a short description of the right breast tissue, pointing out where it is and any potential issues it may have. People often utilize the DDSM dataset to find out how successfully CAD systems and ML algorithms identify BC.
    3. **Insight into BC**: The insight into BC dataset is a free resource of digital mammography pictures for BC. From this dataset, we obtain four images of the left and right breasts, sourced from 100 patients, and displayed in each of the 115 cases. Each image comes with a full description of the identifying breast, including where it is located and any abnormalities including masses, crowds, and structural distortions. The INbreast dataset is widely utilized for assessing the efficacy of computerized systems and machine learning algorithms in breast cancer diagnosis [2].

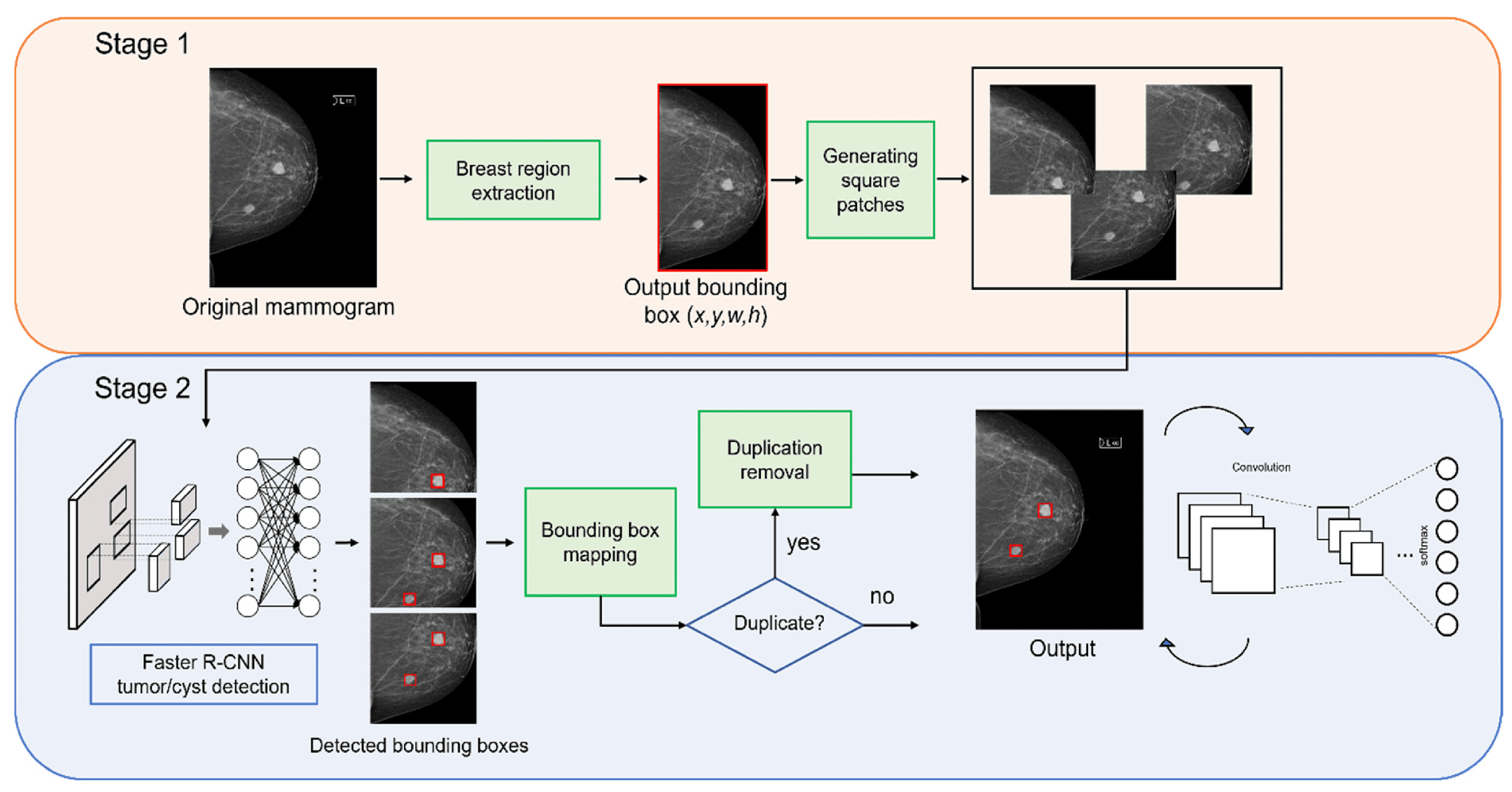
The dataset resources utilized in breast cancer detection research offer a variety of imaging modalities, such as mammography, ultrasound, and histopathology pictures, which function as benchmarks for the development and assessment of AI models. Popular datasets including Mini-MIAS, DDSM, and the INbreast dataset have annotated mammograms with masses, microcalcifications, and other problems that can be used to train machine learning and deep learning algorithms. These datasets differ in size, image resolution, and clinical particulars, facilitating comparative investigation of model performance. Combining results from several imaging methods makes diagnoses more accurate, and using data augmentation techniques makes them even more reliable. Overall, the availability of standardized datasets has greatly improved the creation of computer-aided detection systems. However, there are still key issues to think about when using them in the real world, such as restricted sample variety and possible dataset bias.

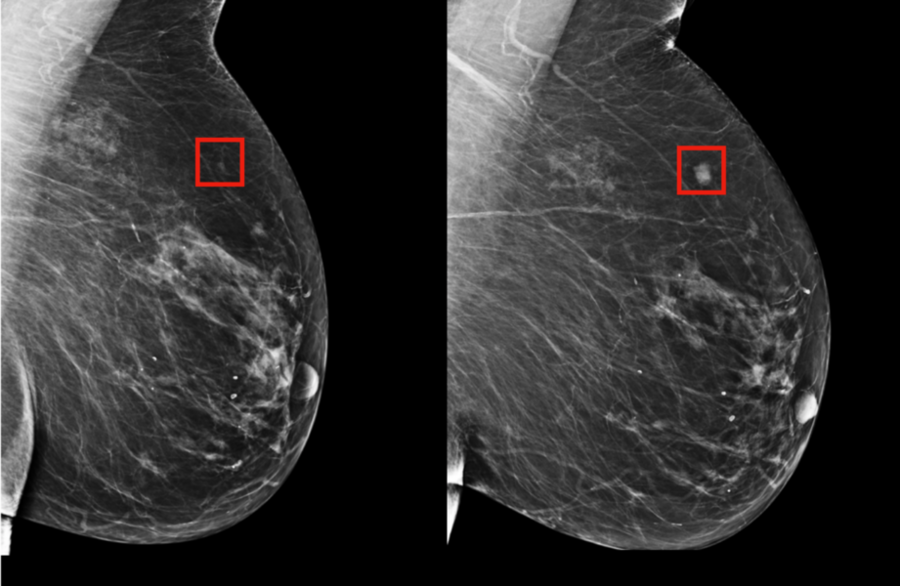
**Representative Images from Breast Cancer Datasets**

The importance of AI and ML in improving breast cancer detection can be better understood through the representative images provided in this section. As shown in Fig. 1, variations in breast density significantly influence the accuracy of cancer detection and remain a critical diagnostic factor. The general workflow of AI-assisted breast cancer detection is illustrated in Fig. 2, where high-resolution mammogram images undergo preprocessing, feature extraction, and classification using convolutional neural networks (CNNs), one of the most widely adopted models for image-based diagnosis. Finally, Fig. 3 demonstrates the output of an AI-assisted system, in which a trained deep learning model accurately highlights the tumor region within the mammogram, thereby supporting radiologists in distinguishing malignant tissues from benign ones.



**FIGURE 1.** Mammogram samples illustrating variations in breast density [3]

 **FIGURE 2.** Process of breast cancer detection using high-resolution mammogram image [4]



**FIGURE 3.** AI-assisted detection that shows the tumor area in a mammography image [5]

**METHODS AND MATERIALS**

This section discusses the most frequently employed approaches for breast cancer detection.

**Convolutional Neural Network (CNN)**

CNNs are commonly employed in medical image analysis for breast cancer screening by autonomously extracting characteristics from mammograms and diverse imaging modalities. This includes preprocessing images, training on labeled datasets, and extracting features in a hierarchy, which lets CNNs find patterns that could mean cancer. After training, these networks can tell the difference between benign and cancerous pictures with high accuracy, which helps radiologists diagnose BC more quickly and accurately. CNNs are very useful in clinical contexts because they can learn from enormous datasets and keep becoming better at what they do.

(f \* g) (t) = (1)

People often use CNNs to classify images, such as identifying BC. A CNN usually has a lot of convolutional layers, pooling layers, and fully connected layers. The CNN technique can be expressed with the following formula:

Output = σ (W \* (Conv2D (Input) + b))

where: - σ represents the activation function (for instance, ReLU or Sigmoid)

- W denotes the weight of the matrix

- b indicates bias

- Conv2 convolutional operation

- Input signifies the input image

**Recurrent Neural Networks (RNN)**

RNNs can be beneficial when assessing sequential information associated with BC, for instance at time-series using static contrast-enhanced MRI or observing changes in over time, yet they are used more rarely in direct medical image than CNNs. By detecting patterns in sequence and capturing temporal dependencies. RNNs in a specific, LSTM network that allows for the assessment of how the lesions progress or respond to treatment over time. RNNs can improve modeling of prediction and assist with clinical decision making by combining imagery data with temporal information, particularly when tracking the course of a disease or accuracy of treatment.

H (t+1) = (2)

Y (t + 1) = (3)

**K-Nearest Neighbors (KNN)**

KNN is an unsupervised ML algorithm used in medical image analysis to diagnose. By sorting photos based on how closely they resemble labeled training samples. This method evaluates each new mammography or breast image against a collection of existing examples within the feature space, where features are typically derived through techniques such as shape descriptors or texture analysis. KNN is very useful when it's important to understand how the categorization process works. It does this by selecting the most common class among the new image's nearest neighbors and giving it that label. KNN can be a good starting point for simple classification jobs and can be useful when there isn't a lot of data, even though it isn't as strong as more complex models like CNNs.

d(p, q) = d(q, p) =

= (4)

KNN technique assigns classes to new data points based on the majority vote from their closest neighbors.

The KNN method can be expressed with the following equation:

Class(X) = argmax (5)

where:

- x represents the new data point that needs to be classified

- k indicates the count of the nearest neighbors

- Yi denotes the class label of the i-th nearest neighbor

- c refers to the class label

- δ (Yi, c) is the Kronecker delta function (1 if Yi is equal to c, otherwise 0)

**Logistic Regression Method**

Logistic Regression is a frequently employed statistical method for binary classification tasks, including the identification of BC.

Formula: The logistic regression model equation (Equation 6):

P(Y=1) = (6)

where: - P(Y=1) is the chance that the positive class will happen   
- e is the base of the natural log  
- z is a linear combination of the model and the input features X.

**Random Forest Method**

Random Forest refers to the random selection of data and training that data to construct a decision tree. It is an ensemble learning technique that combines several decision trees to improve the accuracy and dependability of predictions. The Random Forest method is given by Equation 7.

Prediction(x) = (7)

where: - x represents the input data point, - n denotes decision trees, - which signifies the weight of ith decision tree.

We use DL and ML techniques for detection and classification of BC. Within the field of AI, DL and ML are subsets that focus on applying algorithms to analyze data and derive predictions. A specific kind of ML that employs multi-layered neural networks are called deep neural networks, hence the term "deep “learning. These networks are especially effective for tasks like speech and picture recognition because they can automatically learn characteristics from raw input.

This section showed how widely machine learning and deep learning are used for breast cancer detection. Convolutional Neural Networks (CNNs) are the best and most widely used models for classifying pictures. However, Recurrent Neural Networks (RNNs), K-Nearest Neighbors (KNNs), and Random Forests are other ways to look at sequential patterns, similarity-based categorization, and ensemble decision-making. Each technique individually improves the accuracy of diagnosis, the extraction of features, and the classification of tumors. Researchers and doctors can create successful computer-aided detection systems that improve early diagnosis and help doctors make decisions by using these different methods. The effectiveness of these models relies on the quality of datasets, appropriate preprocessing, and fair training methods to ensure reliable and unbiased outcomes in practical applications.

**APPLICATION OF PROGNOSTIC TOOLS**

In medical imaging, the goal is to make the best picture of the human body possible. In medical imaging, ML and DL are very important for finding and classifying BC cells. CNNs assist find and identify cancers in ultrasound, MRI, and mammography pictures, just like they do for tumors. They can also tell the difference between benign and malignant lesions. Automated segmentation techniques can help define neighboring tissues in imaging examinations and tumor boundaries. This can help ensure that treatment planning and assessment are done without mistakes. ML and DL predictive models evaluate imaging features and patient data to ascertain the probability of breast cancer development or recurrence. By looking into how cancers respond to therapy

**System Architecture for Breast Cancer Prognostic Tools**

Complete system architecture for finding breast cancer has four parts illustrated as follows:

1. *Region of Interest (ROI)*

In breast cancer imaging, regions of interest (ROIs) are specific areas of breast tissue that are set aside for close examination. These ROIs can include lesions that look worrisome, cancers, or areas where the tissue density looks strange. For example, in mammography, ROIs could be found by looking for microcalcifications, masses, or changes in the way tissue is structured. In ultrasound imaging, ROIs may be delineated by the detection of hypoechoic or hyperechoic lesions. By focusing on these specific ROIs, radiologists improve their ability to effectively identify and categorize breast lesions, which helps find them earlier.

*B. Characteristic extrication*

In BC imaging, extracting characteristics means finding and measuring certain attributes or properties of breast cancers or lesions. These properties may include morphological features like shape, size, and edges, as well as textural features like density and regularity. Additional characteristics may encompass kinetic elements, including enhancement patterns and washout rates. Radiologists and computer-aided detection (CAD) systems can tell the difference between benign and malignant tumors by looking at and extracting these properties. This will lead to more accurate BC diagnosis.

*C. Categorization*

Almost all of the models used established and well-developed classifiers to sort breast masses. A standard SVM classifier to tell the difference between aggressive and benign breast cancer. A random forest classifier for dividing mammograms into groups of benign and malignant. A computer-aided detection (CAD) system based on YOLO that uses a fully connected neural network (FCNN) to find breast cancer and classify breast masses.

**TESTING AND DISCUSSION**

In this section, we describe the experimental results from numerous prominent research conducted on a frequently employed dataset and offer discussion based on these findings.

* 1. **Key performance indicators (KPIs)**

Usually, the accuracy (defined in Equation 8), precision (defined in Equation 9), and recall (also known as sensitivity, defined in Equation 10) from the confusion matrix are used to evaluate detection systems. Also, different statistical metrics, such as the F1 score (see Equation 11).

Accuracy = (8)

Precision = (9)

Recall = (10)

F1 score = (11)

Where, the values obtained in the confusion matrix are:

-TP = True positives

-FP = False positives

-TN = True negatives

- FN = False negatives

Similarly, Receiver Operating Characteristics (ROC), and its Area under Curve (AUC) are frequently utilized to evaluate detection outcomes.

**Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC)**

The Receiver Operating Characteristic (ROC) curve is a commonly utilized performance assessment tool in medical image analysis, especially for the identification of breast cancer. It is a graph that shows the trade-off between a classifier's sensitivity and specificity by plotting the True Positive Rate (Sensitivity) versus the False Positive Rate (1 – Specificity) at different threshold levels. A perfect classifier would reach a point in the top-left corner of the ROC space, which would mean that it is very sensitive and has very few false positives.   
The Area Under the Curve (AUC) is a single number that summarizes the ROC curve. Numbers closer to 1.0, mean that the test works well, while numbers around 0.5 mean that the test is random. Models with higher AUC values are deemed more reliable for finding breast cancer since they are better at telling the difference between malignant and benign instances at different decision thresholds.

**AI Biases in Breast Cancer Detection**

AI models for finding breast cancer have been very accurate, but they aren't perfect. AI bias is when predictions are wrong because the training data isn't balanced or full, which makes some groups of patients more likely to be wrong. Imbalance in datasets, lack of representation of certain groups, and technical and acquisition bias are all things that might cause AI bias. Bias can lead to wrong or late diagnoses among groups that aren't well represented, less trust between doctors and patients, and possible legal and moral problems when biased diagnostic tools are used. Some ways to reduce the risk are data diversity, fairness-aware ML models, Explainable AI (XAI), and ongoing validation. AI-powered breast cancer diagnosis can be more accurate, fair, equal, and trustworthy if these problems are solved.

**CONCLUSION**

This paper looked at different machine learning and deep learning algorithms that are commonly used to find and diagnose breast cancer. These algorithms include CNN, RNN, KNN, and Random Forest. The importance of datasets, area of interest (ROI) extraction, feature selection, and performance indicators in building strong diagnostic models was also stressed. The review shows that AI-based early detection systems can help radiologists make better decisions in the clinic and improve the accuracy of diagnoses.   
These AI techniques could change the world, but they can only be trusted if the training data is fair and includes everyone. AI bias, which can happen because of differences in demographics, datasets, or imaging, can make diagnoses less accurate for some groups of patients and lead to unfair healthcare results. This problem needs to be solved in order for it to work in real life. Future research must prioritize the collecting of different datasets, the integration of fairness-aware algorithms, and the deployment of explainable AI to guarantee equitable diagnostic performance. Using this method can improve AI-based breast cancer detection, which will make it more accurate and encourage ethical, unbiased use in the clinic.

**REFERENCES**

1. Y. Lu, J.-Y. Li, Y.-T. Su, and A.-A. Liu, “A review of breast cancer detection in medical images,” 2018 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4, 2018.
2. Y. del C. J. Gaona, “Breast medical images classification through the application of deep learning processing technologies,” M.S. thesis, Universitat Politècnica de València, 2024.
3. S. M. Astley, “Experts developed a deep learning model that can estimate breast density,” Radiology Business, Apr. 10, 2023.
4. Ibrokhimov and J. Y. W. Kang, “Two-stage deep learning method for breast cancer detection using high-resolution mammogram images,” Applied Sciences, vol. 12, no. 9, Art. no. 4616, 2022.
5. Yala, P. G. Mikhael, F. Strand, G. Lin, K. Smith, Y.-L. Wan, L. Lamb, K. Hughes, C. Lehman, and R. Barzilay, “Toward robust mammography-based models for breast cancer risk,” JAMA Oncology, vol. 7, no. 5, pp. 734–741, 2021.
6. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, “SMORE: A self-supervised anti-aliasing and super-resolution algorithm for MRI using deep learning,” IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 805–817, 2021.
7. Mallick, C. R. Behera, B. K. Paikaray, and S. Mishra, “Machine learning approaches with effective feature selection for improving breast cancer prediction,” in Proc. 2023 IEEE 2nd Int. Conf. Ind. Electron.: Developments & Applications (ICIDeA), pp. 157–162, 2023.
8. S. Saxena, S. Shukla, and M. Gyanchandani, “Pre-trained convolutional neural networks as feature extractors for diagnosis of breast cancer using histopathology,” International Journal of Imaging Systems and Technology, vol. 30, no. 3, pp. 239–248, 2020.
9. Y. Liang, D. Schott, Y. Zhang, Z. Wang, H. Nasief, E. Paulson, W. Hall, P. Knechtges, B. Erickson, and X. A. Li, “Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
10. Y. Mo, C. Han, Y. Liu, M. Liu, Z. Shi, J. Lin, B. Zhao, C. Huang, B. Qiu, Y. Cui, L. Wu, X. Pan, Z. Xu, X. Huang, Z. Li, Z. Liu, Y. Wang, and C. Liang, “HoVer-Trans: Anatomy-aware HoVer-Transformer for ROI-free breast cancer diagnosis in ultrasound images,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
11. Saber, M. Sakr, O. M. A. Seida, A. Keshk, and H. Chen, “A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique,” Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 5, pp. 5657–5665, 2021.
12. S. Nanglia, M. Ahmad, F. Khan, and N. Z. Jhanjhi, “An enhanced predictive heterogeneous ensemble model for breast cancer prediction,” Computers, Materials & Continua, vol. 69, no. 1, pp. 885–897, 2022.
13. M. Swain, S. Kisan, J. Chatterjee, M. Supramaniam, S. Mohanty, and N. Jhanjhi, “Hybridized machine learning based fractal analysis techniques for breast cancer classification,” Computers, Materials & Continua, vol. 66, no. 2, pp. 1841–1854, 2020.
14. S. P. Shinde and M. Dixit, “Deep learning approach for breast cancer detection from histopathology images,” South Eastern European Journal of Public Health, vol. 14, pp. 2383–2397, Nov. 2024.
15. M. J. Umer, et al., “A framework of deep learning and selection-based breast cancer detection from histopathology images,” Computational Systems and Engineering, vol. 45, no. 2, pp. 1–14, 2023.
16. M. A. Naji, et al., “Machine learning algorithms for breast cancer prediction and diagnosis,” Procedia Computer Science, vol. 191, pp. 487–492, 2021.
17. V. Nemade and V. Fegade, “Machine learning techniques for breast cancer prediction,” Procedia Computer Science, vol. 218, pp. 1314–1320, 2023.
18. R. Rabiei, et al., “Prediction of breast cancer using machine learning approaches,” Journal of Biomedical Physics & Engineering, vol. 12, no. 3, pp. 297–304, 2022.
19. C. Mallick, et al., “Enhancing breast cancer risk prediction through comprehensive ensemble machine learning analysis: A study on clinical, genetic, and demographic factors,” International Journal of Internet Manufacturing and Services, vol. 11, no. 2, pp. 191–209, 2025.
20. X. Qian, et al., “A multimodal machine learning model for the stratification of breast cancer risk,” Nature Biomedical Engineering, vol. 9, no. 4, pp. 356–370, 2025.
21. L. K. Singh and K. Shrivastava, “An enhanced and efficient approach for feature selection for chronic human disease prediction: A breast cancer study,” Heliyon, vol. 10, no. 5, e03774, 2024.
22. P. B. Dash, H. S. Behera, and M. R. Senapati, “Deep learning based framework for breast cancer mammography classification using ResNet50,” in Computational Intelligence in Pattern Recognition (CIPR 2022), Lecture Notes in Networks and Systems, vol. 480, Springer, Singapore, 2022.
23. C. Mallick, C. R. Behera, B. K. Paikaray, and S. Mishra, “Machine learning approaches with effective feature selection for improving breast cancer prediction,” in Proceedings of the 2023 IEEE 2nd International Conference on Industrial Electronics: Developments & Applications (ICIDeA), pp. 157–162, IEEE, 2023.