Breast Cancer Detection Techniques: A Review

Amar Paul Singh[1], Er. Arvind Bhatia[2], Vineet Rana[3]

[1]Research Scholar Department of Computer Science HPU –Shimla (H.P),
Email: singhamarpaul1@gmail.com
[2]Assistant Professor CGC-Landran Punjab, Email: Arvind.5965@cgc.edu.in
[3]Research Scholar Department of Computer Science HPU –Shimla (H.P),
Email: ranavineet.75@gmail.com

Received: 18.04.2024 Revised : 16.05.2024 Accepted: 24.05.2024

ABSTRACT
Of all malignancies diagnosed in women, breast cancer accounts for one-third. 18% of cancer-related fatalities globally are mostly caused by this. Breast cancer used to be the leading cause of mortality for women, but since 1985, lung cancer has taken over that spot. Due to inadequate knowledge of the subject, this malignancy cannot be prevented. However, cutting-edge breast cancer therapy methods are incredibly effective. Pre-processing, feature extraction, and classification are some of the processes involved in the identification of breast cancer. This research reviews the machine learning architectures for the identification of breast cancer.

Keywords: Breast Cancer, Machine Learning, Deep Learning, Feature Extraction

1. INTRODUCTION

The combined occurrences of the next three top-ranked female cancer kinds in the United States, namely lung and bronchus, colon and rectum, and uterine corpus, are higher than those of breast cancer. This is among the cancer forms that affect women most frequently. Despite the development and widespread use of numerous diagnosis and treatment methods, breast cancer continues to rank among the top two cancer types in both countries in terms of annual fatalities [1]. It is advised to utilize a variety of screening methods to identify breast cancer that is smaller than 2 cm in size because tumour size is one of the key characteristics that is positively connected with long-term mortality. Breast cancer subtypes differ greatly in both their pathological characteristics and the recommended therapies.

1.1 Breast Cancer Detection

Figure 1's illustration of the general architecture of the breast cancer detection model shows five sequential blocks. Every block makes use of the output from the module before it. An original breast picture (such as one obtained from histology, mammography, ultrasound, or MRI) serves as the first module's input. The final output's value captures the characteristics of the tumour in the input image [2].

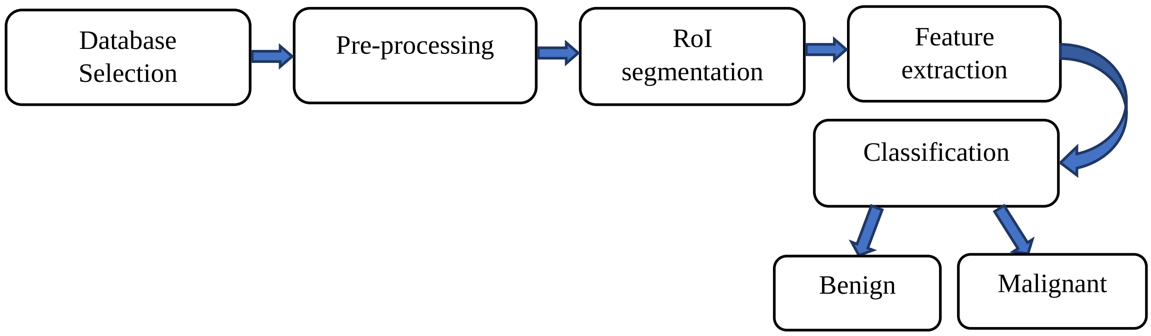

Figure 1. Breast Cancer Detection Framework

All the modules shown in Figure 1 are described as follows.
a. Database selection: In this stage, experimental work is conducted using the breast cancer data set. To diagnose breast cancer, a range of imaging techniques are required, including histology imaging, X-ray (mammography), and ultrasound imaging. The MITOS, MICCAI-AMIDA13, and Camelyon16 histology imaging databases are widely used [3]. Images from a mammogram can be utilised to identify a variety of

suspect abnormalities, such as large and microscopic calcifications. Mini-MIAS, DDSM, and Inbreast are the three main mammography databases.

b. Pre-processing: Imaging artifacts and consistency issues caused by various imaging situations might have a significant impact on subsequent steps. Picture pre-processing approaches must effectively remove variability and artefacts to increase detection accuracy. There are numerous pre-processing methods, including filtering and contrast enhancing. The noise in the raw photos is removed during filtering to minimize its influence on subsequent diagnoses [4]. To determine whether noise is present in a window, the basic idea is to adopt a window sliding of a specific size in the line direction of the image. The mean, variance, and spatial correlation values of each sliding window are then calculated. The pixel values of the particular window are changed to the mean value when noise is discovered. To improve contrast between the suspicious tumour and the tissues around, a contrast augmentation is also used. For this, the original image's histogram is modified to become uniformly distributed. Following this procedure, the image's grey scale increases, improving the contrast and bringing out the finer features in the image [5].

c. RoI (Region of Interest) segmentation: The framework's subsequent module deals with the area's self-directed segmentation.To recover more usable information from a photograph, autonomous digital image segmentation divides a digital photo into multiple fragments. The most similar pixels or regions within a picture are used to segment it. Features can take many different forms, and the approaches employed change depending on the needed attributes. Depending on the categorization criteria employed, different techniques or concepts may be used in segmentation [6]. The picture segmentation module separates the region of interest from the backdrop. Edge-based, threshold-based, and region-based methods are some of the traditional segmentation techniques based on digital image processing.

d. Feature Extraction: The segmented required area which was acquired in the first module, is used to compute and derive features in the third block. These traits are specific values that were taken from a digital imagery. The computed propertiescan shed light on the characteristics of the area of concern from which they were formed. Fractal textures, morphological features, and texture features are the three fundamental feature forms that can be retrieved from the segmented photo [7].

e. Classification: Machine learning algorithms are utilised to categorise tumours as benign or malignant in the fifth and final module. The features that were retrieved in the prior unit are used to determine classification. The performance of the classification architecture is more accurate the more relevant and indicatory the suspicious area in the image is. The classification of cancer images using machine learning algorithms has long been a popular area of study.There are two different research methods: creating neural networks and applying conventional classifiers [8].Neural networkshave the benefit of overlooking blocks during the retrieval of features. To abstract information from an image supplied to a classifier, traditional models need calculations.

1.2 Machine learning Models breast cancer detection

Machine learning (ML) presents an alternate strategy to conventional prediction modelling to address current issues and enhance the reliability of melanoma classifiers. Machine learning methodologies were created as a result of earlier work on pattern recognition and computerized statistical learning. To discover complex correlations among many heterogeneous risk variables, they rely more on software tools and models and less on generalizations. By continually decreasing certain objective functions of events that were anticipated and verified, this is achieved. Models linked to cancer prognosis and survival have been enhanced by ML in regards of the accuracy and validity of predictions [9]. The following is a discussion of some popular machine learning models for detecting breast cancer:

i. Random Forest: Random Forest is among the most popular and efficient algorithms. It consists of several decision trees and creates the class, which is the mode of the class' output, using individual trees. A forest of trees is created once the decision trees are built using the sample sets. It is useful for problems like this classification challenge and others like regression. Regression employs a substantial number of trees in training and then displays the classes or mean forecasts of each individual tree [10]. This method is similar to that previously described. After being trained on distinct subsets of the same dataset, a variety of deep decision trees are aggregated to reduce variation.

ii. Support Vector Machine: SVM is regularly applied to address pattern identification and classification problems when a dataset has exactly 2 classes. SVMs are used to identify the ideal hyperplane that separates the classes. To determine which class it relates to, the classifier employs an input pattern referred to as a feature vector. Although it categorizes data that may be separated linearly, feature vectors may not always be linearly separable. By utilizing the kernel approach, this is resolved. Input data are mapped to higher dimensional data using kernel techniques, and SVM offers a rapid training process. It can be used for regression and pattern categorization [11]. The kernel function chosen affects how well

an SVM classifier performs. Various kernel functions are used for different categorization challenges. In this project, SVM was implemented using the Scikit-Learn SVC class. SVM, however, may require a lot of memory and be challenging to interpret and modify.
iii. Naïve Bayes: One of the best machine learning techniques for binary classification is the naive Bayes algorithm. The Bayes theory and the solid assumption that an attribute's influence on a class is regardless of the values of every other attribute serve as the foundation for this simple probabilistic classifier. The Bayes theorem's equation is: [12]:

$$P(B|A) = \frac{P(B|A)\,P(A)}{P(B)}$$

Class prior probability P(A), predictor prior probability P(B), posterior probability P(A|B), and likelihood P(B|A) are all indicated in equation. The Naive Bayes classifier has the benefit of not being reliant on the volume of training data. It can estimate the classification-related parameters from a minimal set of training data. It functions better in difficult platforms like spam detection, diagnostics, and meteorology. It is suitable when the input has a high degree of dimension.

2. LITERATURE REVIEW

2.1 Breast Cancer Detection using Artificial Intelligence

M. Li, et.al (2021) proposes a computer-based technique called CNN to classify and distinguish the cancer imagery [13]. This work focuses on pre-training two CNNs with various designs, and then utilizing them for automatically extracting the features, integrating them extracted from the two designs, and lastly utilizing the classifier to compute the combined properties. According to the results of the trial, this strategy has an accuracy rate of 89%, making it more accurate than conventional methods in classifying breast cancer photos.
P. E. Jebarani, et.al (2021) anticipated additional metric for assessing the performance of a KMC and a GMM [14]. A combined hybrid form for segmenting and detecting the breast cancer was applied. The proposed strategy is significant to classify the cells as benign and malignant. The simulated outcomes evaluated the efficiency of this strategy for diagnosing the breast cancer at early stage. With greater speed and precision, this method enables medical professionals to identify breast cancer. The multi-variant study and forecast rate for the suggested technique were determined using an ANOVA test.
G. Wadhwa, et.al (2020) developed a deep neural network based upon medical photos to identify melanoma [15]. CNN (Convolutional Neural Network) algorithm was implemented for this purpose. By applying the CNN model DenseNet-201, features are retrieved. The classification issue has two levels: normal and dangerous. The dataset evaluated for the classifier, BreakHis, proved effective by getting recall, accuracy, and precision of 99%, 95.58%, and 90% respectively, and an F1-score of 89%. Testing results and comparisons with similar research help to demonstrate the design's effectiveness and its extremely consistent performance.

Table 1. Breast Cancer Detection using Artificial Intelligence

Author	Year	Technique Used	Findings	Limitations
M. Li, et.al	2021	computer-based technique called Convolutional neural network	According to the results of the trial, this strategy has an accuracy rate of 89%, making it more accurate than conventional methods in classifying melanoma photos.	This technique was not classified 3D data issues and unable to tackle the issue of fusing information
P. E. Jebarani, et.al	2021	Hybrid of K-means and a Gaussian mixture model	The recommended approach is crucial for differentiating between benign and malignant tumours. Medical experts can discover malignancy more promptly and precisely with this technology.	The precision of this technique was mitigated in complicated cases.
G. Wadhwa, et.al	2020	deep learning	A 95.58% accuracy	This technique was

		technique	rate, 0.90 and 0.99 precision and recall rates, and a 0.89 F1-score were achieved using this method.	ineffective to recognize the particular regions for classifying the breast cancer.

2.2 Breast Cancer Prediction using Artificial Intelligence

S. Alghunaim, et.al (2019) created nine models using three reliable algorithms—SVM, DT, and RF—to support the recognition of breast malignancy [16]. In order to demonstrate how the two platforms (Spark and Weka) behave while handling enormous data sets, an experimental comparison of the two was done. The experimentation's authors found that the scaled SVM classifier, which outperformed the other classifiers with an accuracy of 99.68% in the Spark environment, had the best performance with the GE dataset.

Z. Huang, et.al (2022) established a HCRF (Hierarchical Clustering Random Forest) system. In order to execute a clustering analysis on decision trees, the similarity between each DT was assessed using the hierarchical clustering [17]. Exemplary trees were chosen from fragmented groups in order to produce HCRF with low similarity and high accuracy. The VIM method was also used in this study to enhance the chosen feature number for breast cancer prediction. The suggested technology's efficacy was evaluated using a number of metrics. Test results demonstrated that the HCRF algorithm delivered the highest accuracy of 97.05% on the WDBC and 97.76% on WBC dataset. The research's suggested technique proved a useful instrument for finding malignancy.

D. Sun, et.al (2019) developed a MDNNMD for the breast cancer prognosis [18]. The design of the plan of strategy and the synthesis of multi-dimensional data were the peculiar aspects of the methodology. In comparison to prediction strategies using one-dimension data and other current methodologies, the simulation findings demonstrated that the suggested strategy performed preferentially. The original code that TensorFlow 1.0 performed.

Table 2. Breast Cancer Prediction using Artificial Intelligence

Author	Year	Technique Used	Findings	Limitations
S. Alghunaim, et.al	2019	SVM, DT, and RF	The test findings revealed that the scaled SVM classifier with accuracy of an accuracy of 99.68% in the Spark climate beats the other classifiers.	The efficiency to classify the breast cancer was alleviated in case of balanced dataset and feature selection methods.
Z. Huang, et.al	2022	Hierarchical Clustering Random Forest (HCRF)	Test results demonstrated that the HCRF algorithm delivered the highest accuracy of 97.05% on the WDBC and 97.76% on WBC dataset.	The structural diversity was not analyzed and the relevant constraints were not optimized in this model.
D. Sun, et.al	2019	MDNNMD	In comparison to prediction strategies using one-dimension data and other current methodologies, the simulation findings demonstrated that the suggested strategy performed preferentially.	This technique makes the deployment of multi-dimensional data for recognizing the survival time of patients suffered from the breast cancer, which creates issues for researches in case of scarcity of data.

2.3 Breast Cancer Prediction or Detection using Artificial Intelligence

S. Y. Siddiqui, et.al (2021) presented a cloud-based IoMT model for breast cancer prediction [19]. The suggested architecture was used to identify the stages of breast cancer. The experiment's findings revealed that the training and validation phases' accuracy was 98.86% and 97.81%, respectively. Furthermore, they showed accuracies of 99.69% for ductal carcinoma detection, 99.32% for lobular carcinoma, 98.96% for mucinous carcinoma, and 99.32% for papillary cancer. The impact of the proposed deep learning-enhanced intelligent prediction of breast cancer stages demonstrated more accuracy than current classic approaches, showing its ability to reduce the breast cancer incidence rate.

U. Naseem, et.al (2022) suggested a methodology using an ensemble of classifiers to distinguish and forecast breast cancer. Diverse ML methods were employed first, followed by an ensemble of different ML algorithms [20]. This article then provided a summary of ML algorithms and a group of several classifiers, for automatically diagnosing BC and determining its forecast. On two standardized datasets, this paper also presented and compared numerous ensemble architectures and other variations of tried ML-supportedarchitectures. This research also concentrated on how balanced class weighting affected the prognostic dataset and evaluated how well it performed in comparison to other studies. The findings indicated that the ensemble procedure outperformed other existent approaches and obtained accuracy of 98.83%.

S. Lee, et.al (2019) suggested an EAL methodology in order to detect the breast cancer [21]. After obtaining the shrinking uncertainty from the classification and prediction of the breast cancer using the current ML approaches on histopathology pictures, the AL was applied to forecast the malignancy. The trial findings indicated that the recommended methodology performed more effectively in comparison with others while detecting the breast cancer.

Table 3. Breast Cancer Prediction or Detection using Artificial Intelligence

Author	Year	Technique Used	Findings	Limitations
S. Y. Siddiqui, et.al	2021	IoMT cloud-based model	The experiment's findings revealed that the training and validation phases' accuracy was 98.86% and 97.81%, respectively.	The IPBCS-DL model results in maximizing the computing complicated nature of the mechanism.
U. Naseem, et.al	2022	ANN and ensemble of various classifiers	The findings indicated that the ensemble procedure outperformed other existent approaches and obtained accuracy of 98.83%.	The proposed framework consumed much cost and becomes complicated to be implemented on an enormous sized data.
S. Lee, et.al	2019	EAL (ensemble based active learning) technique	The trial findings indicated that the recommended methodology performed more effectively in comparison with others while detecting the breast cancer.	This technique was unable to analyze the entire slide image.

CONCLUSION

The first indicators of beginning, non-palpable breast cancer are calcifications. Calcifications typically have a connection to ductal carcinoma in situ (DCIS). But invasive tumours can also develop calcification. In the course of breast imaging therapy, it is regarded as the primary indicator of malignant tumours. In screening programmes, calcifications are portrayed as the only sign of cancer in about 41.2% of women. The models of machine learning are not efficient as compared to deep learning. The performance of major models is analysed in view of accuracy, precision and recall.

REFERENCES

[1] Poorti Sahni, Neetu Mittal, "Breast Cancer Detection Using Image Processing Techniques", 2019, Advances in Interdisciplinary Engineering, pp. 813-823

[2] M. Naderan and Y. Zaychenko, "Convolutional Autoencoder Application for Breast Cancer Classification," 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), 2020, pp. 1-4

[3] F. Rahman, T. Mehejabin, S. Yeasmin and M. Sarkar, "A Comprehensive Study of Machine Learning Approach on Cytological Data for Early Breast Cancer Detection," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6

[4] N. Khuriwal and N. Mishra, "Breast Cancer Detection from Histopathological Images Using Deep Learning," 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), 2018, pp. 1-4

[5] S. Kaya and M. Yağanoğlu, "An Example of Performance Comparison of Supervised Machine Learning Algorithms Before and After PCA and LDA Application: Breast Cancer Detection," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-6

[6] E. S. Wahyuni, R. P. Rasmi and S. Murnani, "Performance Improvement of Breast Cancer Diagnosis based on Mammogram Images using Feature Extraction and Classification Methods," 2021 IEEE International Biomedical Instrumentation and Technology Conference (IBITeC), 2021, pp. 77-82

[7] Mengfan Li, "Research on the Detection Method of Breast Cancer Deep Convolutional Neural Network Based on Computer Aid", 2021, IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)

[8] Than Than Htay, Su Su Maung, "Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image", 2018, 18th International Symposium on Communications and Information Technologies (ISCIT)

[9] Mehrnaz Ronagh, Mohammad Eshghi, "Hybrid Genetic Algorithm and Particle Swarm Optimization Based Microwave Tomography for Breast Cancer Detection", 2019, IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)

[10] NirmineHammouch, Hassan Ammor, "Microwave Imaging for Early Breast Cancer Detection using CMI", 2019, 7th Mediterranean Congress of Telecommunications (CMT)

[11] Monica Ezzat Gamil, Mariam Mohamed Fouad, Mohamed A. Abd El Ghany, Klaus Hoffinan, "Fully automated CADx for early breast cancer detection using image processing and machine learning", 2018, 30th International Conference on Microelectronics (ICM)

[12] Tanny Chavez, Nagma Vohra, Jingxian Wu, Magda El-Shenawee, Keith Bailey, "Spatial Image Segmentation for Breast Cancer Detection in Terahertz Imaging", 2020, IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting

[13] M. Li, "Research on the Detection Method of Breast Cancer Deep Convolutional Neural Network Based on Computer Aid," 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 536-540

[14] P. E. Jebarani, N. Umadevi, H. Dang and M. Pomplun, "A Novel Hybrid K-Means and GMM Machine Learning Model for Breast Cancer Detection," in IEEE Access, vol. 9, pp. 146153-146162, 2021

[15] G. Wadhwa and A. Kaur, "A Deep CNN Technique for Detection of Breast Cancer Using Histopathology Images," 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), 2020, pp. 179-185

[16] S. Alghunaim and H. H. Al-Baity, "On the Scalability of Machine-Learning Algorithms for Breast Cancer Prediction in Big Data Context," in IEEE Access, vol. 7, pp. 91535-91546, 2019

[17] Z. Huang and D. Chen, "A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering Random Forest Algorithm," in IEEE Access, vol. 10, pp. 3284-3293, 2022

[18] D. Sun, M. Wang and A. Li, "A Multimodal Deep Neural Network for Human Breast Cancer Prognosis Prediction by Integrating Multi-Dimensional Data," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 3, pp. 841-850, 1 May-June 2019

[19] S. Y. Siddiqui et al., "IoMT Cloud-Based Intelligent Prediction of Breast Cancer Stages Empowered With Deep Learning," in IEEE Access, vol. 9, pp. 146478-146491, 2021

[20] U. Naseem et al., "An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers," in IEEE Access, vol. 10, pp. 78242-78252, 2022

[21] S. Lee, M. Amgad, M. Masoud, R. Subramanian, D. Gutman and L. Cooper, "An Ensemble-based Active Learning for Breast Cancer Classification," 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019, pp. 2549-2553