Breast Cancer Detection Using Machine Learning

RITIK RANJAN
Amity Institute of Information and Technology
Amity University, Noida, INDIA

Abstract: - Drawing meaningful information from images is a crucial aspect of image processing. It is extensively utilised in several medicinal fields. Early diagnosis and treatment results, particularly in the case of cancer detection, can be greatly impacted by timely picture analysis. In view of the disease's rising global occurrence, this study investigates the diagnosis and treatment of breast cancer. Two important modalities—MRIs and mammograms—are used in the proposed approach to increase the accuracy of tumour identification. Numerous segmentation techniques, such as edge detection and threshold methods, are used to separate the tumorous regions. Several operators are also applied to the resulting images, and quantitative validation is performed using entropy measurements. Healthcare providers might potentially save lives and heal patients faster if breast cancer is discovered early on.

Keywords: image preprocessing, image, breast cancer, segmentation, mammogram, histogram analysis entropy, mri.

Received: March 9, 2024. Revised: August 8, 2024. Accepted: September 2, 2024. Published: October 11, 2024.

1. Introduction

Image processing encompasses a range of techniques aimed at extracting meaningful data from images, constituting a fundamental aspect of medical imaging for diagnosing and treating diseases. The process typically involves image acquisition, enhancement, and segmentation. Image enhancement, crucial for improving image characteristics, aims to adjust image attributes for specific tasks and observers, whether human or machine. Image segmentation, on the other hand, classifies or identifies sub-patterns within an image, facilitating the partitioning of the image into segments with similar attributes. Segmentation methods can be broadly categorized into those based on discontinuity and similarity, each offering distinct approaches to image segmentation.

Breast cancer, characterized by the abnormal growth of cells in breast tissues, remains a significant cause of mortality among women worldwide. Recognizing the signs and symptoms of breast cancer is imperative for timely intervention, although their presence does not necessarily confirm the presence of the disease. Diagnostic procedures such as mammograms, utilizing X-rays to detect breast abnormalities, are crucial for early detection. Subsequent tests, including ultrasound and MRI, provide clearer images for further assessment, with biopsy serving as the definitive diagnostic procedure for confirming cancer presence. Understanding the signs, symptoms, and diagnostic procedures is essential for effective breast cancer management. Breast cancer remains a significant public health concern worldwide, with early detection being pivotal for successful treatment and patient outcomes. Over the past decade, the integration of machine learning (ML) techniques

2. Literature Review

Breast cancer detection has shown promising results, offering opportunities for more accurate and efficient diagnostic tools. In this literature review, we explore recent studies and advancements in utilizing ML algorithms for breast cancer detection, focusing on their methodologies, performance metrics, and implications for clinical practice.

2.1 Traditional Methods vs. Machine Learning Approaches

Traditionally, breast cancer detection relied on techniques such as mammography, ultrasound, and biopsy. While these methods have been effective, they often suffer from limitations such as subjective interpretation and variability in results. ML algorithms, on the other hand, have demonstrated the ability to analyze complex patterns within medical imaging data, leading to improvements in accuracy and efficiency.

2.2 Imaging Modalities and Feature Extraction

Studies have utilized various imaging modalities, including mammograms, magnetic resonance imaging (MRI), and digital breast tomosynthesis (DBT), to capture detailed information about breast tissue characteristics. ML algorithms are then employed to extract features from these images, such as texture, shape, and intensity, which are subsequently used for classification and prediction tasks.

2.3 Classification Algorithms in Breast Cancer Detection

Several ML algorithms have been applied to breast cancer detection, including support vector machines (SVM), artificial neural networks (ANN), random forests, and convolutional neural networks (CNN). These algorithms leverage labeled datasets to learn patterns indicative of malignant or benign tissue, enabling accurate classification of breast lesions.

2.4 Performance Evaluation and Validation

The performance of ML models in breast cancer detection is assessed using metrics such as sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC-ROC). Cross-validation techniques are commonly employed to evaluate model generalization across different datasets, ensuring robustness and reliability.

2.5 Challenges and Future Directions

Despite the advancements made in ML-based breast cancer detection, several challenges remain. These include the need for large and diverse datasets, interpretability of ML models, and integration into clinical workflows. Future research directions may focus on enhancing the explainability of ML algorithms, incorporating multi-modal data fusion, and deploying automated decision support systems in real-world healthcare settings.

3. Proposed Methodology

To enhance the accuracy of breast cancer detection, the following methodology is proposed:

3.1 Image Conversion

Converting images to grayscale is a vital preprocessing step, eliminating the complexities associated with color images and facilitating subsequent analysis.

3.2 Image Filtering

Image filtering techniques, such as Gaussian smoothing filters, are employed to remove noise and enhance image clarity, enabling more effective analysis.

3.3 Histogram Analysis

Histogram analysis provides insights into the tonal distribution of digital images, aiding in the characterization of image properties and guiding further processing steps

3.4 Edge Detection Based Segmentation

Edge detection methodologies identify abrupt changes in intensity within an image, facilitating the segmentation of regions of interest. Various edge detection operators are applied to extract relevant features from mammograms and MRIs.

Steps for edge detection method.

Input: An image under consideration (MRI or mammogram)

Output: Magnitude of the slope of the image

Step 1: Provide the input image.

Step 2: Utilize edge detection theorems to determine gradients using:

- Horizontal mask (Gx)
- Vertical mask (Gy)

Step 3: Apply a mask to the input image.

Step 4: Generate distinct images for the horizontal and vertical masks.

Step 5: Combine the outcomes to ascertain the magnitude of the gradient.

Step 6: Output the gradient magnitude of the image, representing the slope.

3.5 Thresholding-based Segmentation

Thresholding techniques categorize image pixels based on their intensity values, simplifying the segmentation process and delineating regions of interest.

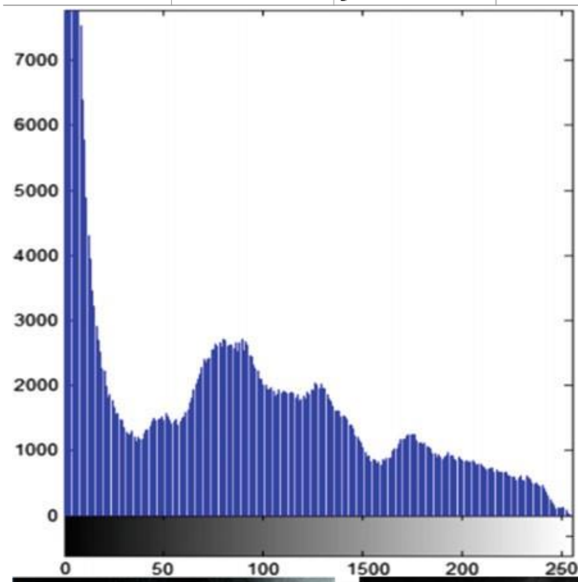
If $P_{ij} \leq T$, Otherwise, the pixel is allocated to category 2 (Table 1).

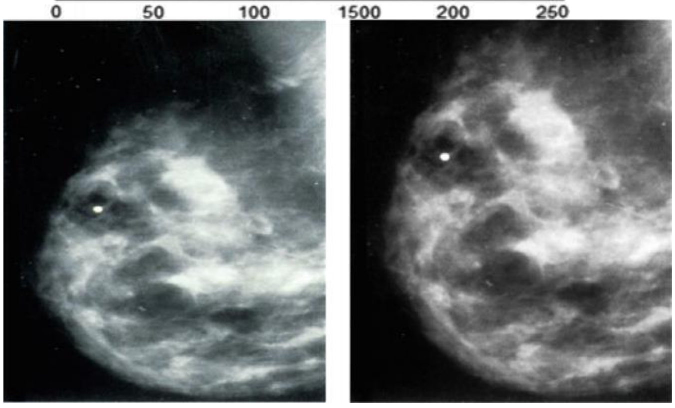
4. Performance Assessment

Performance evaluation is conducted using entropy, a statistical measure of randomness, to quantify the information content of processed images.

Table 1 Comparison between thresholding- and edge-based techniques

Segmentation method	Explanation	Strength	Weakness
Thresholding technique	Threshold values can be computed on the basis of histogram peaks	No prior information is required	Spatial details missing
Edge-based technique	Based on detection of discontinuity	Images displaying difference among different objects have good results	Not preferable for images with high number of edges





(a) Original (b) Grayscale

Fig. 1. Mammogram image

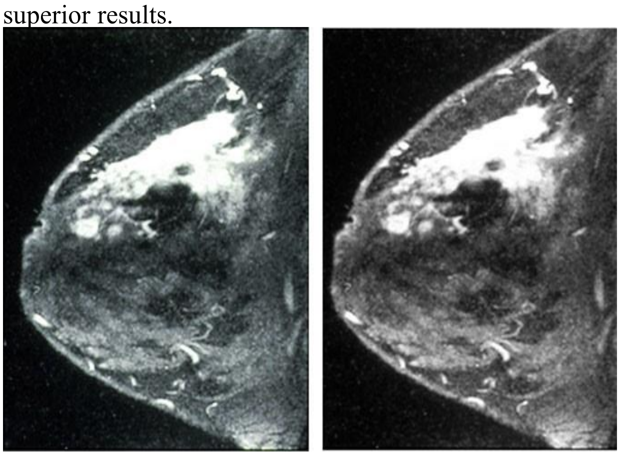
Image filtering, segmentation, and edge detection techniques are assessed based on their ability to enhance image quality and accurately identify tumorous regions.

5. Simulation and Results

In the presented research, the process involves applying a Gaussian filter to mammogram and MRI images to effectively remove the noise depicted in Figures 1 through 6. Specifically, Figure 1 displays the initial mammogram alongside its grayscale representations, while Figure 2 illustrates the original MRI image and its grayscale conversion. Additionally, histogram analysis of the original mammogram and MRI images is depicted in Figures 3 and 5, respectively.

Histogram analysis of the original mammogram and MRI images shown in Figs. 3 and 5 respectively.

The proposed methodology is applied to mammograms and MRIs, resulting in the successful extraction of tumorous regions. Comparative analysis of segmentation techniques reveals the efficacy of edge detection methods over thresholding, with specific operators yielding superior results.



(a) Original (b) Grayscale

Fig. 2. MRI image

comparative analysis has been done by measuring the entropy of output images (Figs. 7, 8, 9, 10, and 11).

According to Table 2, the entropy values for edge detection and thresholding methods applied to MRI images are 0.0987 and 0.0763, respectively. For mammogram images, the corresponding entropy values are 0.1002 and 0.040. The comparison indicates that the edge detection method yields superior results compared to thresholding for both types of images.

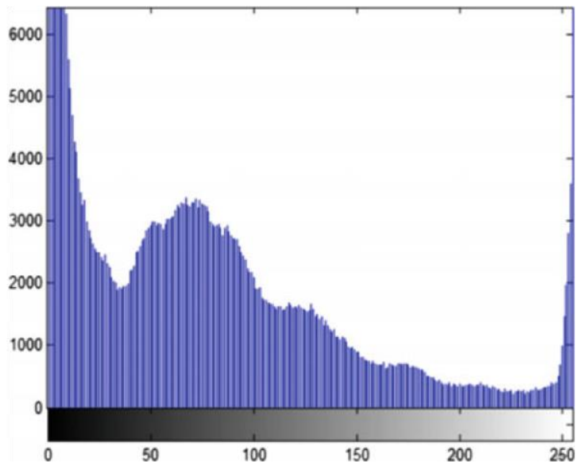
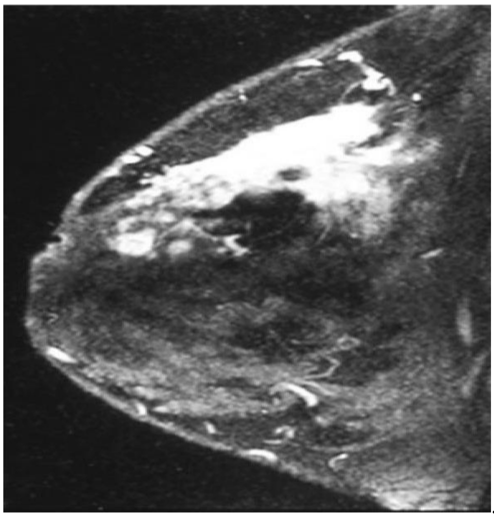


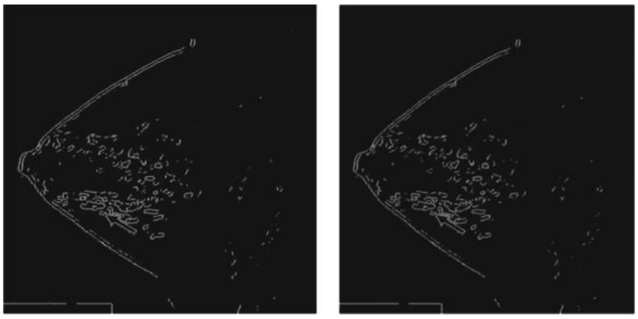
Fig. 3. Histogram of original MRI

Measuring parameter	Technique	Mammogram	MRI
Entropy	Edge detection	0.1002	0.0987
	Thresholding	0.0401	0.0763

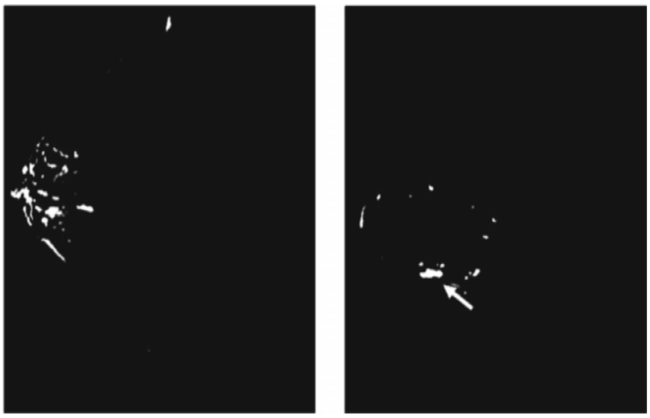
Table 1Comparative analysis of entropy among mammogram and MRI images



Fig, 4. Filtered image of MRI



(a)Thresholding for MRI (b) Thresholding for Mammogram



(a) Thresholding for MRI (b) Thresholding for Mammogram

Fig. 5. (a) Edge detection for MR (b) Edge detection for Mammogra

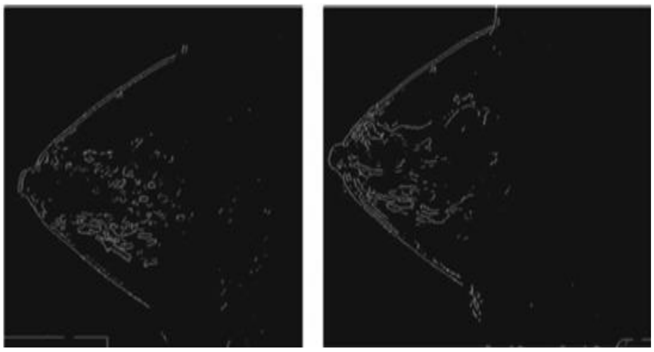


Fig. 6. (a)Thresholding for MRI (b) Thresholding for Mammogram

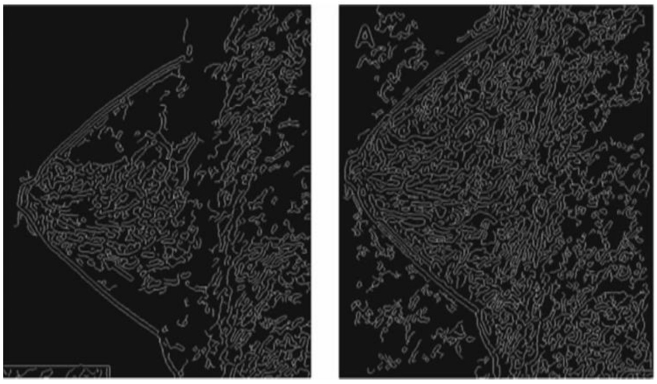
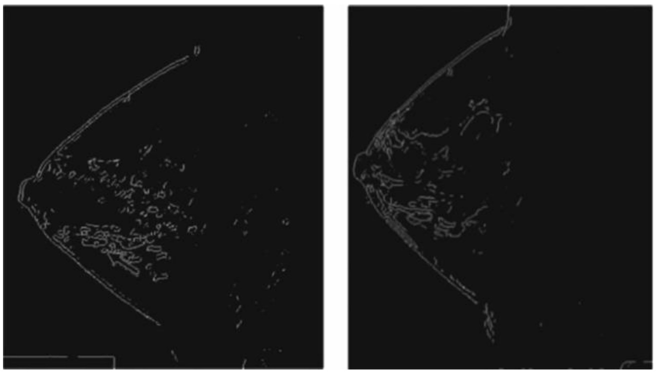
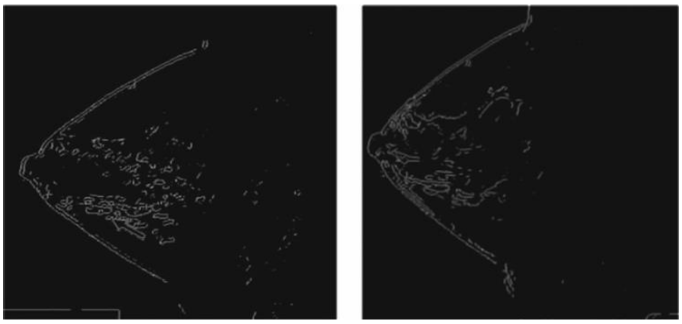


Fig. 7. (a)Thresholding for MRI (b) Thresholding for Mammogram



(a)Mammogram (b)MRI

Fig. 8 a,b Sobel operator for Mammogram and MRI



(a) Mammogram (b)MRI

Fig. 9. a b Canny operator on Mammogram and MRI

Edge detection operator	Mammogram	MRI
Sobel	0.1024	0.1010
Canny	0.3707	0.5334
Prewitt	0.1031	0.1012

Table 3 presents a comparative analysis of edge detection operators applied to both mammogram and MRI images. According to the analysis, the Canny edge detector achieves the highest entropy value, indicating its superiority over other edge detection methods.

Edge detection operators and its resultant images shown in Fig. 9

Table 3 presents a comparative analysis of edge detection operators applied to both mammogram and MRI images. According to the analysis, the Canny edge detector achieves the highest entropy value, indicating its superiority over other edge detection methods.

6. Conclusions

In conclusion, the proposed methodology demonstrates promising results in enhancing image quality and accurately identifying tumorous regions in breast cancer detection. Edge detection algorithms, particularly the Canny operator, exhibit superior performance compared to thresholding methods, underscoring their potential for improving diagnostic accuracy in breast cancer imaging.

References

[1] Merida-Gubern A, M Kallenberg, RM Mann, Marti R, Karssemeijer N (2015) Breast density and segmentation estimation in breast MRI..

[2] R Ramani, NS Vanitha (2013) The pre-processing techniques for breast cancer detection inmammography images.

[3]Al- MS Tarawneh (2012) Detection of Lung cancer using image processing techniques.

[4] SU Mageswarim, M Sridevic (2013) An experimental analysis and study of different image segmentation techniques.

[5] S Kaymaka, A Helwan, D Uzuna (2017) Breast cancer classification of image using artificialneural networks

[6] Breast Cancer Cases in India (2016) Indian Council For Medical Research, New Delhi. https://icmr.nic.in/guide/cancer/Breast_Cancer.pdf

[7] AK Singh, B Gupta (2015) A novel approach for breast cancer segmentation and detection ina mammogram.

[8] The Breast Cancer Record https://www.who.int/news-room/fact-sheets/detail/breast-cancer.

[9] Haq A. U., “Detection of Breast Cancer Through Clinical Data Using Unsupervised Supervised Feature Selection Techniques,”
[10] Khuriwal. N and Mishra N., “Breast cancer diagnosis using voting ensemble machine learning algorithm,”
[11] S. Ara, Das A. and Dey A., “Benign and Malignant Breast Cancer Classification using Machine Learning Algorithms”
[12] Abed B. M. , “A hybrid classification approach algorithm for breast cancer diagnosis”
[13] Khuriwal N. and Mishra. N, “Breast Cancer Diagnosis Using Learning Algorithm”
[14] Prakash and Visakha K., “Breast Cancer Malignancy Prediction Using Learning Neural Networks,”
[15] Telsang V. A. and Hegde K., “Breast Cancer Analysis Prediction using Machine Learning Algorithms.
[16] Mojrian S. et al., “Machine Learning Hybrid Model of Learning Extreme Machine Radial basis function for Breast Cancer Diagnosis and Detection; a Multilayer Fuzzy Expert System,”