Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Cancer Guard: Early Detection of Breast Cancer

Shaista Khanam[1], Soham Kavathkar[2], Manali Bhadange[3], Pawan Singh[4]

[1]Profesor, Department of Electronics and Telecommunication, Vidyavardhini's College of Engineering and Management, University of Mumbai, Mumbai, India.

[2]BE Student, Department of Electronics and Telecommunication, Vidyavardhini's College of Engineering and Management, University of Mumbai, Mumbai, India.

E-mail: [2]Soham.200181105@vcet.edu.in, [3]Manali.s210021203@vcet.edu.in, [4]Pawan.200431101@vcet.edu.in

Abstract

Breast cancer stands as the most prevalent form of cancer among women, globally contributing to the highest number of cancer-related deaths. Timely detection of abnormalities significantly enhances the prospects of successful treatment and reduces mortality rates. Hence an automatic detection will be very useful for medical practitioner. This research introduces a novel framework for enhancing breast cancer detection and stage classification by integrating image processing techniques such as Gray Level Co-occurrence Matrix (GLCM) and Convolutional Neural Network (CNN) techniques. Initially, mammographic images undergo preprocessing to improve quality, followed by GLCM feature extraction for capturing textural information. With the help of GLCM technique, accuracy of the network can be increased by extracting various features. A CNN model is then employed for automatic feature learning and classification. This framework enhances the accuracy of distinguishing between malignant and benign tissues and extends to stage detection, enabling classification into various stages. Experimental results demonstrate the effectiveness of the proposed approach in achieving high precision and recall rates, suggesting potential for clinical integration to improve patient outcomes and streamline healthcare workflows.

DOI: https://doi.org/10.36548/jscp.2024.1.008
Received: 02.03.2024, received in revised form: 07.04.2024, accepted: 21.04.2024, published: 02.05.2024

Keywords: Breast Cancer Detection, Deep Learning, Convolutional Neural Networks, GLCM, Mammography, Malignant & Benign.

1. Introduction

Breast cancer is widely recognized as one of the most prevalent types of cancer and ranks as the second leading cause of death among women worldwide. In 2002, it was the second most diagnosed cancer globally, with over one million new cases reported. Despite improvements in early detection and understanding the molecular underpinnings of breast cancer biology, nearly 30% of patients diagnosed with "early-stage" breast cancer experience disease recurrence. Detecting breast cancer early is crucial for effective treatment. Imaging plays a central role in diagnosis, with techniques like MRI, SPECT, mammography, CT, and PET being utilized to identify and monitor patients across different stages of the disease.[1].

Mammography, a widely adopted method, utilizes low-dose X-rays for breast examination. It is straightforward and economically accessible[11]. Cancerous masses and calcium deposits appear brighter on mammograms. Presently, mammography stands as the gold standard for early-stage breast cancer detection, identifying lesions before they become palpable clinically [2].

Deep learning algorithms are pivotal in the early detection of diseases, facilitating timely intervention. Machine learning techniques, particularly those leveraging deep learning methodologies, surpass conventional approaches by employing sophisticated algorithms to identify crucial regions of interest in digital mammograms[14-15]. Advanced methods such as Grey-Level Co-occurrence Matrix (GLCM) and Image Segmentation are utilized to extract features such as texture and shape from mammographic images [7-10]. This process enhances tumour visibility, enabling more precise detection of abnormalities in breast tissue [4]. The GLCM technique in breast cancer texture analysis helps identify specific features distinguishing normal and abnormal tissues, aiding in early detection and treatment decisions. Researchers explore GLCM's effectiveness compared to other methods and its correlation with clinical parameters for improved diagnosis and patient outcomes [6].

A Convolutional Neural Network (CNN) is a specialized type of multi-layer feedforward neural network primarily designed for classification tasks, particularly in image classification [11-13]. Each neuron within a CNN receives inputs from a local receptive field in the previous layer, enabling it to extract local features. The convolutional layers consist of multiple feature maps, with neurons arranged in planes to capture different aspects of the input data. CNNs have revolutionized deep learning, particularly in image processing, by enhancing the effectiveness of deep learning algorithms [3].

2. Related Work

The previous work which has been done is revealed in this section, for feature selection and extraction in mammogram images.

Vaira Suganthi Gnanasekaran, Sutha Joypaul, Parvathy Meenakshi Sundaram, and Durga Devi [2] presents a deep learning-based algorithm for breast mass classification in mammograms. their research highlights the effectiveness of the deep learning algorithm in accurately classifying breast masses from mammograms. The algorithm likely demonstrates high sensitivity and specificity, indicating its potential for aiding radiologists in detecting and diagnosing breast cancer at an early stage. Moreover, the study may discuss the algorithm's performance compared to traditional methods, showcasing its superiority in terms of accuracy and efficiency. Overall, their findings contribute to the advancement of computer-aided diagnosis systems for breast cancer detection, offering a promising approach to improve clinical outcomes and patient care.

Mahesh S. Kedare and Dr. V. B. Kamble [4] focuses on utilizing 2D mammography images and the Gray-Level Co-occurrence Matrix (GLCM) for breast cancer detection. Their study involves extracting features from mammography images using GLCM and employing a classification algorithm to differentiate between cancerous and non-cancerous regions. Through their investigation, they present findings indicating the potential enhancement in breast cancer detection accuracy, thereby suggesting promising implications for clinical applications.

Tongjai Yampaka, Duangjai Noolek introduces [5] a novel method for early breast cancer staging through the integration of mammography and biopsy data. It presents findings regarding the development of a predictive model for accurate staging and the identification of significant features contributing to improved accuracy. This research offers promise for enhancing breast cancer diagnosis and treatment through personalized approaches based on integrated data.

Athraa H. Farhan and Mohammed Y. Kamil [6] researchers investigated texture analysis methods - Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gray-Level Co-occurrence Matrix (GLCM) - for breast cancer detection. They found that LBP effectively captured local texture patterns, while HOG complemented it by providing gradient and spatial orientation information. GLCM, on the other hand, quantified spatial relationships between pixel intensities. Combining these techniques enabled robust detection and classification of breast cancer lesions, offering potential for improved diagnostic accuracy and clinical decision-making.

3. Proposed Work

The suggested method in Figure 1, uses a combination of Image Processing as well as GLCM, for extracting various features and these extracted features are used for classifying the mammogram images into malignant and begin, which are often referred as cancerous and non-cancerous respectively.

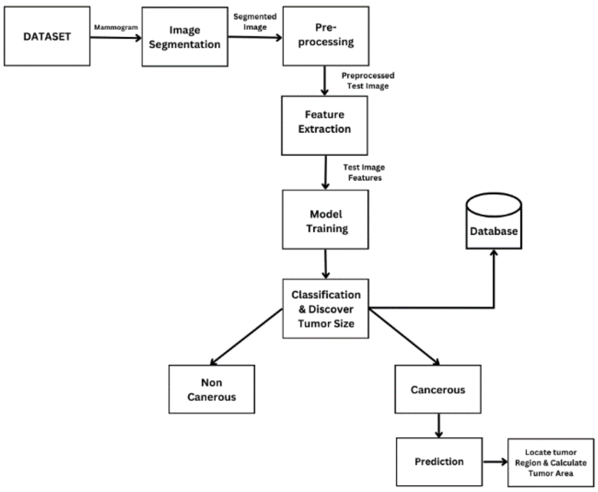

Figure 1. Block Diagram for Breast Cancer Detection.

A. Dataset

The dataset contains Breast Mammogram Images with the labels of:

1. Density levels in breast imaging are categorized as follows: A: (1) "Almost entirely fatty" denotes breasts predominantly composed of fat, found in approximately 1 in 10 women. B: (2) "Scattered areas of fibro glandular density" indicate some dispersed density, with the majority being fatty tissue, seen in about 4 in 10 women. C: (3) "Heterogeneously dense" signifies a mixture of dense and fatty tissue, with dense areas being predominant, observed in around 4 in 10 women. D: (4) "Extremely dense" describes breasts with nearly all tissue being dense, occurring in about 1 in 10 women.

2. Tumor: Figure 2, represents mammogram image for Benign (Non-Cancerous) & Figure 3, represents mammogram image for Malignant (Cancerous).

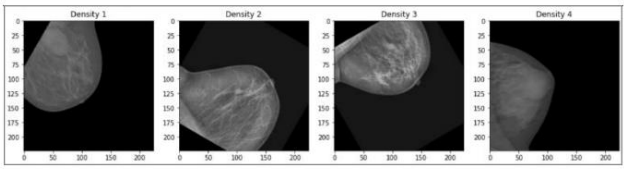

Figure 2. Mammogram Images for Benign.

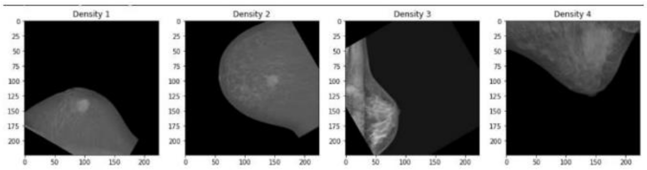

Figure 3. Mammogram Images for Malignant.

B. Image Segmentation

Image segmentation is a fundamental process in image analysis, where an image is partitioned into separate regions, each representing a distinct object or feature. This technique serves critical roles in numerous applications, including object recognition, feature extraction, and image processing. Moreover, image segmentation significantly contributes to advancing robotic scene understanding and decision-making in various domains such as medicine, computer vision, and technology. The various methods used for image segmentation are as follows:

i. Grayscale Conversion

Grayscale conversion simplifies complex color images into a single channel, aiding image segmentation by highlighting brightness changes. This conversion also ensures compatibility with segmentation algorithms designed for grayscale images. In Python, RGB to grayscale transformation is achieved using the cv2.cvtColor function, particularly with the cv2.COLOR_BGR2GRAY parameter.

ii. Histogram Equalization

Histogram equalization is a widely used image enhancement technique in cancer diagnosis, particularly in medical imaging such as mammograms. It enhances contrast and visual detail in images, thereby making small abnormalities more visible. By redistributing pixel usage in the image, histogram equalization aids radiologists and machine learning algorithms in clearly identifying tumors or abnormalities.

iii. Bilateral Filters

Bilateral filters are pivotal in enhancing medical images such as mammograms for breast cancer screening. By improving image clarity and reducing noise, they facilitate the identification of abnormalities and tumors while preserving vital structural details. This preservation of key features leads to more accurate diagnoses and better patient outcomes. Specifically, in the enhancement process, bilateral filtering is applied to grayscale images through functions like cv2.bilateralFilter, effectively reducing noise while maintaining edge integrity.

iv. CLAHE (Contrast Limited Adaptive Histogram Equalization)

CLAHE is an essential imaging tool in diagnosing breast cancer, particularly in enhancing mammograms for improved diagnostic outcomes. It enhances contrast and visibility, enabling radiologists and machine learning algorithms to identify abnormalities and tumors more accurately. This tool is crucial in improving the effectiveness of breast cancer diagnosis by ensuring clearer and more precise visualization of relevant features in mammographic images.

v. Edge Detection

Sobel Filter- The Sobel filter method enhances edges and structural regions in medical images like mammograms, aiding in the detection of potential tumors by highlighting their boundaries. It improves visibility and emphasizes critical details, contributing to more precise and timely identification of abnormalities in cancer diagnosis.

Canny Filter- The Canny filter is instrumental in identifying breast cancer in medical images such as mammograms. It excels in detecting changes in pixel density, effectively highlighting critical features like tumor borders and patterns in breast tissue. Its application significantly contributes to the early-stage detection of cancer, thereby enhancing the accuracy of tumor diagnosis and classification.

vi. Dilation

Dilation, a common technique in breast cancer analysis, involves morphological operations in image analysis aimed at expanding regions of interest. It benefits by merging and connecting adjacent structures and enhances the detection of potential tumors in medical images, notably mammograms. Its application significantly improves the accuracy and efficiency of breast cancer diagnosis by refining the delineation and visibility of relevant structures within the images.

vii. Adaptive Thresholding

Adaptive thresholding, an imaging technique in cancer diagnosis, enhances accuracy by distinguishing healthy and abnormal tissue based on local characteristics. This method is crucial in early cancer detection, particularly in mammography, aiding precise clinical diagnoses.

viii. Image Pre-processing

Figure 4 represents, images before and after pre-processing. Image Processing is essential for enhancing the quality and visibility of mammogram images to ensure accurate analysis. The blurriness and fuzziness inherent in mammograms, caused by the imaging process and breast structure, require preprocessing. This step is crucial for refining image clarity and contrast, thereby creating optimal conditions for thorough analysis.

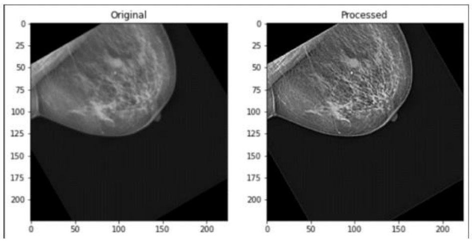

Figure 4. Images before and after Pre-Processing

C. Feature Extraction

Feature extraction in mammography isolates pertinent information from images and patient data to detect abnormalities, focusing on tumor texture and shape. The Gray Level Co-occurrence Matrix (GLCM) method is key for texture extraction, enhancing image segmentation by identifying regions with similar textures and improving boundary distinction. Extracted GLCM features include Contrast, Homogeneity, Dissimilarity, Energy, and Correlation.

D. Model Training

Training models for breast cancer diagnosis entails teaching machine learning or deep learning algorithms to recognize patterns in data, often derived from mammograms or patient records. This process begins with a dataset, where the model adjusts its parameters to minimize prediction errors and enhance its capacity to identify relevant patterns within the provided data.

E. Classification and Detection of Tumor Size

In the classification process, cancer is initially categorized as either cancerous or non-cancerous. If the tumor features, such as size and texture, indicate malignancy, the prediction corresponds to a cancerous diagnosis. Conversely, if the features suggest non-cancerous characteristics, the prediction aligns with a non-cancerous diagnosis.

4. Results and Discussion

The result of breast cancer detection integrates feature extraction, tumour size analysis, and stage detection through the combined utilization of Convolutional Neural Networks (CNN) and Gray Level Co-occurrence Matrix (GLCM) techniques. Through CNN, intricate patterns and features are discerned from mammographic images, allowing for the identification of cancerous abnormalities with heightened accuracy. Simultaneously, GLCM facilitates texture analysis, enhancing the characterization of tumour attributes such as shape, texture, and spatial relationships within the images. By synergizing CNN's ability to capture detailed features and GLCM's texture analysis, the classification model achieves a comprehensive understanding of breast cancer manifestations. Additionally, the integration of tumour size information aids in determining the cancer stage, crucial for treatment planning and prognosis assessment. This amalgamation of techniques enables a holistic approach to breast cancer diagnosis, ensuring precise identification, staging, and subsequent management decisions, ultimately contributing to improved patient outcomes.

1. Grayscale Conversion

Figure 5 represents, Grayscale Conversion technique.

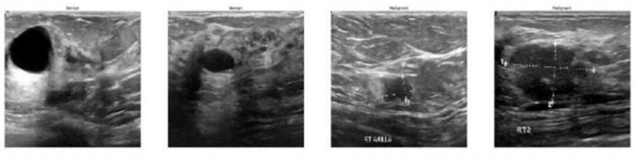

Figure 5. Grayscale Conversion

2. Histogram Equalization

Figure 6, represents Histogram Equalization.

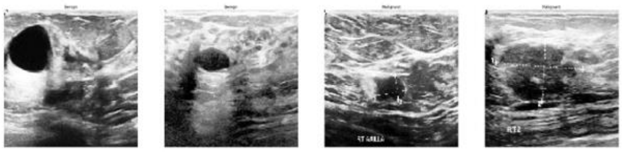

Figure 6. Histogram Equalization

3. Bilateral Filter

Figure 7, represents Bilateral Filter.

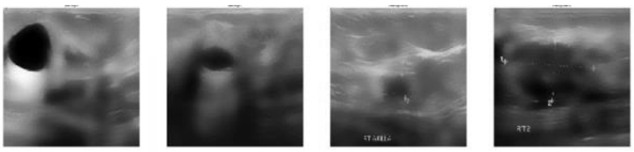

Figure 7. Bilateral Filter

4. **CLAHE (Contrast Limited Adaptive Histogram Equalization)**

Figure 8, represents CLAHE.

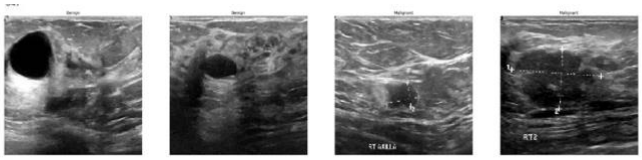

Figure 8. CLACHE

5. **Edge Detection**

i. **Sobel Filter**

Figure 9, represents Edge Detection by using Sobel Filter.

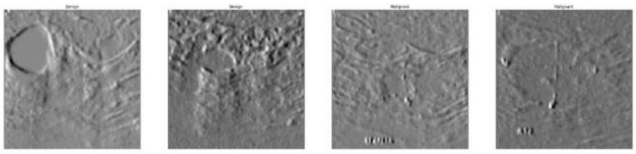

Figure 9. Sobel Filter

ii. **Canny Filter**

Figure 10, represents Edge Detection using Canny Filter.

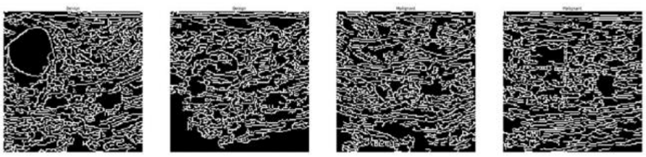

Figure 10. Canny Filter

6. Dilation

Figure 11, represents Dilation.

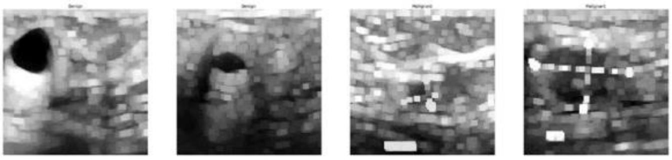

Figure 11. Dilation

7. Adaptive Threshold

Figure 12, represents Adaptive Threshold.

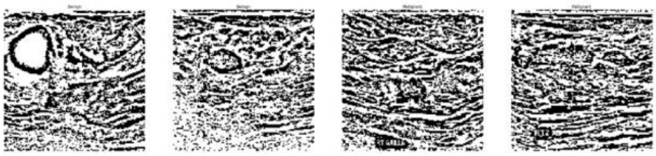

Figure 12. Adaptive Threshold

8. Gray-Level Co-occurrence Matrix

Figure 13, represents Gray-Level Co-occurrence Matrix.

```
GLCM properties for Image_1.jpg: {'contrast': array([[ 40.91680013,  96.53859921,  80.4830237 , 104.00193046],
       [105.11850467,  96.53859921, 187.59256355, 104.00193046],
       [162.29444085, 205.1438195 , 264.791633  , 218.68407597]]), 'dissimilarity': array([[2.74003043, 4.23990026, 3.8849695
7, 4.39936456],
       [4.37843871, 4.23990026, 5.96044482, 4.39936456],
       [5.45046865, 6.2714877 , 7.14344295, 6.43892541]]), 'homogeneity': array([[0.69679672, 0.67848973, 0.68345994, 0.67744
055],
       [0.67881284, 0.67848973, 0.66783842, 0.67744055],
       [0.67110291, 0.66428248, 0.66009317, 0.66342064]]), 'energy': array([[0.62980417, 0.62674134, 0.62755388, 0.62671828],
       [0.62847733, 0.62674134, 0.62383211, 0.62671828],
       [0.62768058, 0.62310143, 0.62035036, 0.6225348 ]]), 'correlation': array([[0.98923997, 0.97467988, 0.97890044, 0.97272
239],
       [0.97236312, 0.97467988, 0.95095005, 0.97272239],
       [0.95733324, 0.94634964, 0.93094717, 0.94280847]])}
GLCM properties for Image_10.jpg: {'contrast': array([[ 42.79818626,  80.75768666,  49.01985906,  72.59705604],
       [114.66606338,  80.75768666, 127.46221445,  72.59705604],
       [179.58680107, 188.99166058, 191.40976487, 157.5988759 ]]), 'dissimilarity': array([[2.53241111, 3.44155322, 2.8017296
6, 3.40600052],
       [4.14285714, 3.44155322, 4.52628298, 3.40600052],
```

Figure 13. Gray-Level Co-occurrence Matrix

9. Convolutional Neural Network

Figure 14, represents Convolutional Neural Network.

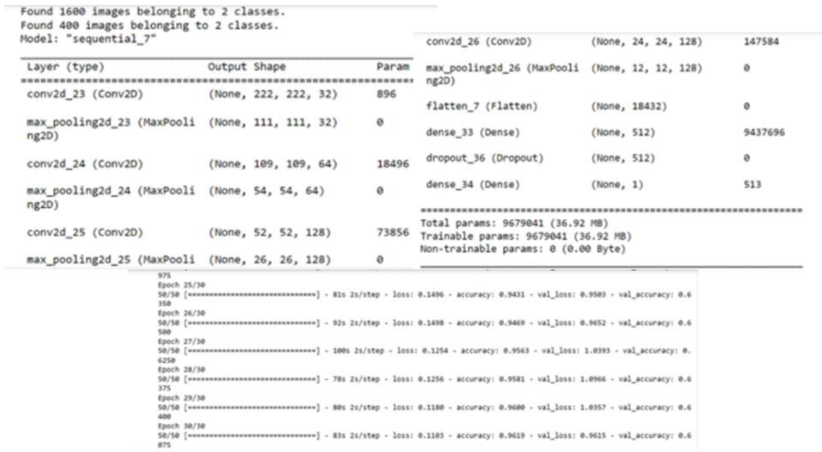

Figure 14. CNN

The table given blow, represented as Table 1, indicates the accuracy of the system with respect to GLCM with Neural Networks and Convolutional Neural Network.

Table 1. Accuracy w.r.t GLCM with Neural Network & CNN

	GLCM with Neural Network	Convolutional Neural Network
Accuracy	70%	96%

10. Stage Detection

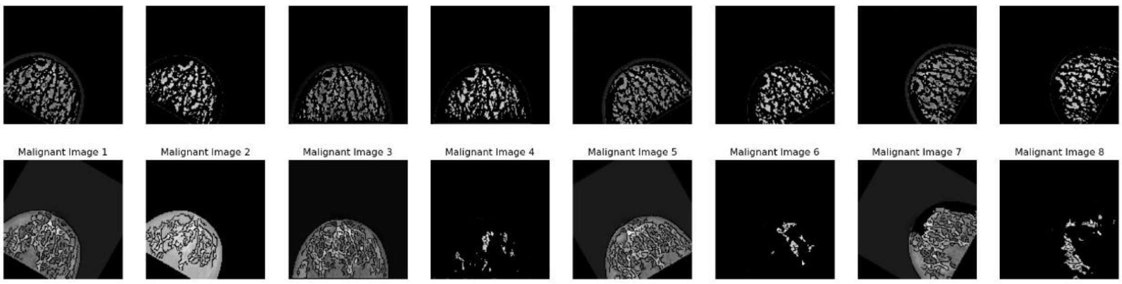

Figure 15. Stage Detection

To detect stages, malignant mammogram image is first preprocessed to reduce noise and enhance features. This involves various image segmentation techniques which includes canny edge detection, inversion, and region filling. With the help of this techniques, tumor size is obtained. Canny edge detection algorithm is then applied to the preprocessed images, this helps to identify edges which may corresponds to the boundary of potential tumors. To locate breast tumors, first, invert the detected edges to make tumors appear brighter. Then, the enclosed areas are filled to identify complete tumor regions. Next, segmentation of these regions is done to isolate potential tumor areas using techniques like thresholding or region growing. Finally, to estimate the size of the segmented regions is done by measuring their dimensions, such as area or perimeter. From the above Figure 15, the mammogram images are classified into low, medium, and high stages with respect to the factor area. Breast having less area of tumor of about 2cm are classified as Low Stage. Mammogram images 4 and 6 represents Low Stage. For medium stage, the area of tumor is between 2 to 5cm. The mammogram image 8 have area between 2 to 5 cm. Hence, it is represented as Medium Stage. Lastly, Breast having tumor size of more than 5cm are of High Stage. Mammogram images 1, 2, 3, 5 and 7 represents High Stage of Breast Cancer.

5. Conclusion

In conclusion, the integration of feature extraction, image processing, model training, classification, and tumour size analysis, facilitated by Convolutional Neural Networks (CNN) and Gray Level Co-occurrence Matrix (GLCM) techniques, presents a comprehensive approach to breast cancer detection and staging. Leveraging CNN, intricate features are extracted from mammographic images, while GLCM aids in texture analysis, enhancing the characterization of tumour attributes. Through image processing techniques, noise reduction and enhancement are achieved, ensuring the quality and clarity of input data for model training. Model training involves optimizing parameters to minimize prediction errors and optimize pattern recognition. Classification, based on extracted features, enables the differentiation between cancerous and non-cancerous tissue. Additionally, tumour size analysis and stage detection contribute crucial information for treatment planning and prognosis assessment. By amalgamating these components, a holistic understanding of breast cancer manifestations is attained, facilitating precise diagnosis and informed treatment decisions. This integrated approach signifies a significant advancement in breast cancer detection, promising improvements in patient outcomes and healthcare delivery.

References

[1] Abunasser, Basem S., Mohammed Rasheed J. AL-Hiealy, Ihab S. Zaqout, and Samy S. Abu-Naser. "Breast cancer detection and classification using deep learning Xception algorithm." International Journal of Advanced Computer Science and Applications 13, no. 7 (2022).pp 223-228

[2] Gnanasekaran, Vaira Suganthi, Sutha Joypaul, Parvathy Meenakshi Sundaram, and Durga Devi Chairman. "Deep learning algorithm for breast masses classification in mammograms." IET Image Processing 14, no. 12 (2020): 2860-2868.

[3] Wang, Zhiqiong, Mo Li, Huaxia Wang, Hanyu Jiang, Yudong Yao, Hao Zhang, and Junchang Xin. "Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features." IEEE Access 7 (2019): 105146-105158.

[4] Mahesh S. Kedare, Dr. V. B. Kamble Mastography Classification using 2 D and GLCM for Detection of Breast Cancer International Journal of Engineering Research & Technology (IJERT) 10(12) 2021. 404 -407

[5] Yampaka, Tongjai, and Duangjai Noolek. "Data Driven for Early Breast Cancer Staging using Integrated Mammography and Biopsy." Asian Pacific Journal of Cancer Prevention: APJCP 22, no. 12 (2021): 4069.2-15

[6] Farhan, Athraa H., and Mohammed Y. Kamil. "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques." In IOP conference series: materials science and engineering, vol. 928, no. 7, p. 072098. IOP Publishing, 2020.1-10

[7] Girija, O. K., and Sudheep Elayiodm. "M. Hybrid method of local binary pattern and classification tree for early breast cancer detection by mammogram classification." Int. J. Recent Technol. Eng 8 (2019): 139-145

[8] .Rathi, Megha, and Vikas Pareek. "Hybrid approach to predict breast cancer using machine learning techniques." International Journal of Computer Science Engineering 5, no. 3 (2016): 125-136.

[9] Fakoor, Rasool, Faisal Ladhak, Azade Nazi, and Manfred Huber. "Using deep learning to enhance cancer diagnosis and classification." In Proceedings of the international conference on machine learning, vol. 28, pp. 3937-3949. 2013.

[10] Loukil, Zainab, Qublai Khan Ali Mirza, Will Sayers, and Irfan Awan. "A Deep Learning based Scalable and Adaptive Feature Extraction Framework for Medical Images." Information Systems Frontiers (2023): 1-27.

[11] Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N. N., & Madabhushi, A. (2017). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Scientific reports, 7, 46450

[12] Heath, Michael, Kevin Bowyer, Daniel Kopans, P. Kegelmeyer Jr, Richard Moore, Kyong Chang, and S. Munishkumaran. "Current status of the digital database for screening mammography." In Digital Mammography: Nijmegen, 1998, pp. 457-460. Dordrecht: Springer Netherlands, 1998.

[13] Karpathy, A. (n.d.). CS231n Convolutional Neural Networks for Visual Recognition. Retrieved May 03, 2016, from = http://cs231n.github.io/convolutionalnetworks/

[14] Kuo, Chung-Feng Jeffrey, Hsuan-Yu Chen, Jagadish Barman, Kai-Hsiung Ko, and Hsian-He Hsu. "Complete, Fully Automatic Detection and Classification of Benign and Malignant Breast Tumors Based on CT Images Using Artificial Intelligent and Image Processing." Journal of Clinical Medicine 12, no. 4 (2023): 1582.

[15] Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., ... & Martí, R. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics, 22(4), 1218-1226

Author's biography

Ms. Shaista Khanam, Assistant professor in Department of Electronics and Telecommunications Engineering, in Vidyavardhini's College of Engineering and Technology, Mumbai University. Author has published more than 13 papers in National and international conference. Area of specializations are Embedded system, IOT, Image processing and deep learning.

Ms. Manali Bhadange, a student in the Department of Electronics and Telecommunications Engineering at Vidyavardhini's College of Engineering and Technology, Mumbai University

Mr. Soham Kavathkar, a student in the Department of Electronics and Telecommunications Engineering at Vidyavardhini's College of Engineering and Technology, Mumbai University.

Mr. Pawan Singh, a student in the Department of Electronics and Telecommunications Engineering at Vidyavardhini's College of Engineering and Technology, Mumbai University