iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VII **Month of publication:** July 2024

DOI: https://doi.org/10.22214/ijraset.2024.63736

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

Harnessing Deep Learning for Accurate Detection of Breast Cancer in Histopathological Imagery

Dhanikonda Ratna Bhavani[1], Ashok Kumar Manda[2], Dr. R. Shyamala Gowri[3], M. Krishnaveni[4], G. Krishnaveni[5], Rahul Ravi[6], Akshatha Naik[7]

[1, 2]CSE Department, [3, 5]Microbiology Department, [4]Biochemistry Department, [6]BSH Department,
[1]NRI Institute of Technology, [2]Vikas College of Engineering and Technology, [3,4,5]Nadar Saraswathi College of Arts & Science,
[6]SCMS Schools Engineering and Technology, [7]Moodlakatte Institute of Technology

Abstract: ***Breast cancer, the most common cancer among women after skin cancer, significantly contributes to the rising mortality rate. Screening mammography is an effective method for detecting masses and abnormalities related to breast cancer. Digital mammograms are especially useful for early cancer detection in asymptomatic women and diagnosing cancer in women with symptoms such as lumps or nipple discharge, thereby reducing mortality and increasing survival rates. Clinicians often face time constraints that can lead to medical errors and incorrect diagnoses due to insufficient time to review patient history thoroughly.***

Implementing machine learning in breast cancer diagnosis enhances accuracy, reduces misclassifications, and saves diagnostic time. This study focuses on the automatic classification of mammogram images into benign, malignant, and normal categories using various machine-learning algorithms. The pre-processed images are classified with the help of a Convolutional Neural Network, providing accurate categorization into benign, malignant, and normal mammograms.

Keywords: Breast Cancer Detection, CNN, benign, Malignant, ROI

I. INTRODUCTION

Breast Asia, the world's largest continent, is home to about 60% of the global population. Breast cancer is the most prevalent type of cancer and the second leading cause of cancer-related deaths among women in Asia, accounting for 39% of all breast cancers diagnosed worldwide (World Health Organization, 2021). The incidence of breast cancer in Asia varies widely across the continent and is generally lower than in Western countries. However, the proportional contribution of Asia to global breast cancer rates is rapidly increasing alongside socioeconomic development. The mortality-to-incidence ratio for breast cancer in Asia is notably higher than in Western countries, primarily because most Asian countries are low- and middle-income countries (LMICs) where breast cancer tends to present at a younger age and more advanced stage, resulting in higher mortality rates (Ginsburg et al., 2021). Diagnostic, treatment, and palliative care services for breast cancer are often inadequate in many Asian LMICs, further contributing to the higher mortality rates (Sung et al., 2021).

According to the World Health Organization (WHO), breast cancer is the most prevalent cancer among women globally, with incidence rates ranging from 19.3 per 100,000 women in Eastern Africa to 89.7 per 100,000 women in Western Europe (WHO, 2021). This high variability is attributed to differences in lifestyle, urbanization, and healthcare access (Bray et al., 2018). Early diagnosis is more feasible in developed countries, while it remains less common in underdeveloped regions, indicating that preventive measures alone are insufficient. Mammography is a standard screening method that can identify suspicious areas in the breast, which are then biopsied to determine whether they are benign or malignant (Duffy et al., 2019). Tissue samples obtained during a biopsy are used to create stained histology slides, which pathologists traditionally examine under a microscope for final diagnosis, staging, and grading (Elmore et al., 2015).

In this context, automatic image analysis and machine learning techniques play a crucial role in enhancing diagnostic accuracy [26]. Recent studies have compared various nuclei segmentation algorithms to classify cases as benign or malignant (Saha et al., 2018). Deep Convolutional Neural Networks (CNNs) have shown remarkable efficiency in visual perception and localization tasks, leading to their adoption in biomedical applications such as breast cancer diagnosis and classification (Litjens et al., 2017). These advancements in machine learning and image analysis hold promise for improving breast cancer detection and outcomes in Asia and beyond

©IJRASET: All Rights are Reserved

II. LITERATURE REVIEW

Paul et al. (2020) developed a Relative Entropy Maximized Scale Space for mobile phone segmentation using morphological opening and closing techniques, complemented by a side-retaining filter to ensure accuracy. Beura et al. (2017) employed traditional cropping methods to select regions of interest (ROI) in mammograms. Beevi et al. (2019) proposed a segmentation method using a Krill Herd Algorithm-based localized active contour model to differentiate cell nuclei from the background, followed by a multiclass classifier based on a deep belief network to categorize cells as mitotic or non-mitotic. Hu et al. (2018) utilized adaptive thresholding segmentation on a multiresolution representation of mammogram images to detect suspicious lesions. Kozegar et al. (2016) introduced a two-stage segmentation method for mass detection in 3D automated breast ultrasound images, starting with an adaptive region growing algorithm based on the Gaussian mixture model (GMM) for a rough boundary estimate, followed by a geometric edge-based deformable model for more precise segmentation.

In feature extraction, Al-Ayyoub et al. (2018) applied a Fuzzy C-Means algorithm based on a Single Pass to extract features from mammographic images and proposed using GPUs to accelerate the process. Albarqouni et al. (2016) introduced a multi-scale CNN AggNet to learn features from crowd-annotated data. Xing et al. (2015) proposed a novel nucleus segmentation method using a deep convolutional neural network (CNN) and a selection-based sparse structure model. Jiang et al. (2016) developed a BCDCNN for breast cancer detection in mammograms, demonstrating the feasibility of CNNs in this domain. Castro et al. (2017) adapted the CNN architecture into a Fully Convolutional Network (FCN) for full mammogram mass detection. Shell et al. (2018) presented a multi-tiered backpropagation neural network (BNN) structure with six neural networks, using four to determine malignant or benign classifications. Carneiro et al. (2015) demonstrated the use of high-level deep learning features for mammogram classification and segmentation map generation.

Lin et al. (2017) proposed a novel framework based on fully convolutional networks for feature learning, reconstructing dense predictions to ensure detection accuracy. Elmoufidi et al. (2019) utilized multiple-instance learning (MIL) algorithms for feature extraction, while Hu et al. (2019) combined the Hidden Markov Tree (HMT) model with the Dual-Tree Complex Wavelet Transform (DTCWT) to extract features from ROI areas for microcalcification detection in mammograms. For classification, Elmoufidi et al. (2019) employed a standard SVM classifier to distinguish between malignant and benign breast cancers. Beura et al. (2017) used a random forest classifier for benign-malignant mammogram classification. Al-many et al. (2018) proposed a YOLO-based computer-aided detection (CAD) system for breast cancer detection, using a fully connected neural network (FCNN) for breast mass classification. A novel early neural network based on transfer learning named 'EARLYNET' was devised and built in this research to automate breast cancer prediction and distinguish benign breast tumors from malignant ones [24] A spatial attention-based neural architecture search network (SANAS-Net) technique that incorporates a spatial attention mechanism, enabling the model to learn and prioritize key regions within mammograms (MMs) [25]

III. METHODOLOGY

A. *Project Modules and Methods for Breast Cancer Detection*

The generalized device structure for breast cancer detection comprises the following key parts. Flow chart mammogram detection and classification is shown in Figure 1.

1) *Image Preprocessing:* Imaging artifacts and inconsistencies due to varying imaging conditions can significantly impact detection accuracy. It is essential to remove variability and artifacts through preprocessing methods to enhance detection performance. Techniques such as noise reduction, image enhancement, and normalization are typically employed to achieve this. In the proposed system, the input image undergoes preprocessing to reduce noise and standardize image size. This step includes smoothing and filtering to enhance image quality and resize the image for consistent analysis.
2) *Segmentation:* Segmentation is crucial to cancer detection as it simplifies the image, making it easier to identify and analyze. Techniques like thresholding, edge detection, and region growing are commonly used to segment the image into more manageable and meaningful parts.
3) *Region of Interest (ROI) Area Segmentation:* Since only specific areas of the whole slide image are relevant for detection, extracting these areas before applying detection techniques is crucial. ROI segmentation helps focus on the most pertinent components of the image, improving the efficiency and accuracy of subsequent analysis. Dhungel et al. (2015) proposed a cascade of deep learning and random forest classifiers to identify masses in mammograms. They employed a morphological scale-space approach for cell segmentation, improving the precision of ROI detection.

©IJRASET: All Rights are Reserved

4) *Feature Extraction:* Raw image data typically has high dimensionality, making it challenging to use directly for classification. Feature extraction involves mapping raw image data into a lower-dimensional feature space which is more relevant to the classification task. This step is critical for capturing essential characteristics of the image, such as shape, texture, and intensity. The proposed method extracts various features based on shape, size, color, intensity, contrast, edges, corners, and regions of interest. This investigation focused on detecting and calculating features in mammograms using Convolutional Neural Networks (CNNs) to distinguish between normal and abnormal mammograms. The use of morphological operations for segmentation significantly improved the classification results. Consequently, the implementation demonstrated that images could be reliably analyzed and classified for further evaluation, capturing the essential characteristics of the image and improving the efficiency of the classification process. Common techniques include texture analysis, histogram of oriented gradients (HOG), and principal component analysis (PCA).
5) ***Classification:*** The extracted features are fed into one or more classifiers to determine whether the ROI areas indicate cancer. Advanced machine learning and deep learning algorithms are often utilized to improve classification accuracy. The extracted features are then classified using machine learning algorithms such as support vector machines (SVM), random forests, or deep learning models like convolutional neural networks (CNN). These classifiers help distinguish between benign and malignant regions with high accuracy.

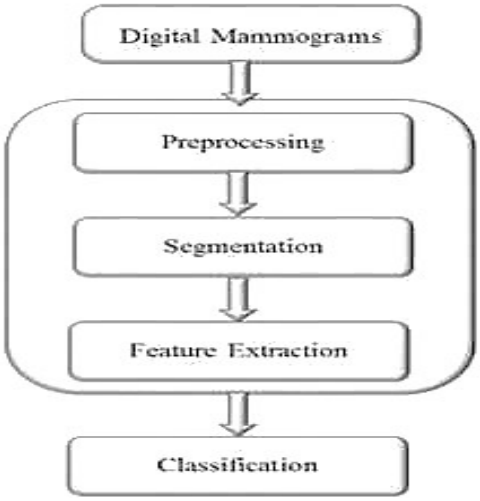


Figure 1. Workflow for Mammogram Analysis and Categorization

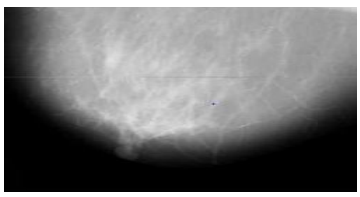
IV. RESULTS AND DISCUSSIONS

A. *Database and Preprocessing Details*

The study utilizes a database from the Mammographic Image Analysis Society (MIAS), which offers an extensive collection of digital mammograms. The MIAS dataset is highly regarded in breast cancer research, containing 161 pairs of mammographic images in Portable Gray Map (PGM) format. These images encompass both normal and pathological cases, making the dataset ideal for developing and evaluating breast cancer detection algorithms.

B. *Image Acquisition*

The dataset's original images, depicted in Figure 2, are grayscale mammograms with a resolution of 1024 x 1024 pixels. These high-resolution images capture detailed structures and potential abnormalities within the breast tissue, providing a rich source of data for analysis.


Figure 2. Input Original image

The original grayscale mammogram image showcases the detailed structure of breast tissue, including any visible abnormalities.

C. Preprocessing Steps

1) *Conversion to Black and White:* Initially, the grayscale mammograms are converted into black and white images, as shown in Figure 3. This conversion is crucial for simplifying the image data and emphasizing the regions of interest (ROI) for further analysis. Thresholding techniques are used to separate the foreground (potential abnormalities) from the background (normal tissue).

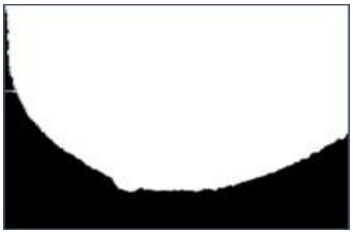


Figure 3. Black and white image

The processed black-and-white image highlights the selected regions of interest for further analysis. By converting images to black and white and applying these preprocessing steps, we ensure that the data fed into machine learning models is clean, consistent, and focused on relevant areas. This preprocessing enhances the detection algorithms' performance, thereby improving the overall accuracy of breast cancer diagnosis.

2) *Noise Reduction:* Digital mammograms may contain noise due to various imaging artifacts and inconsistencies. To improve image quality, noise reduction techniques such as Gaussian smoothing or median filtering are applied. These methods help eliminate unwanted variations, enhancing the clarity of regions that will be processed further. To reduce noise in the images, a Wiener filter is applied. This adaptive filter works to minimize the mean square error between the original and estimated images, effectively reducing noise while preserving important details. The background of the original image is removed, isolating the regions of interest by eliminating empty rows and columns from the black-and-white image. The resultant image, shown in Figure 4

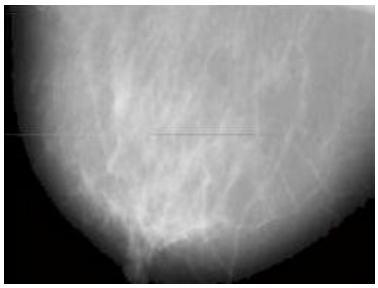


Figure 4. Removing background image

The Wiener filter is employed to reduce noise in an image by using statistical techniques to estimate the desired signal. It operates by minimizing the mean square error between the estimated and actual signals, effectively filtering out unwanted noise. This filter takes into account both the degradation function and the statistical properties of the noise to produce a clearer image. As a result, it is particularly effective in enhancing image quality where the noise characteristics are well-understood. The Weiner filter is used to reduce the noise in an image as shown in figure 5.

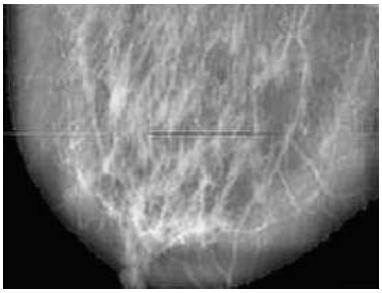


Figure 5. Weiner filter Image

3) *Normalization:* To ensure consistent brightness and contrast across all images in the dataset, image normalization is performed. This standardization facilitates more accurate feature extraction and classification by providing uniform input data.
4) *Resizing:* For computational efficiency and compatibility with machine learning models, the images are resized to a uniform dimension, typically 256 x 256 pixels. This resizing balances the need for detail with processing power constraints.
5) *ROI Selection:* Post-preprocessing, the black and white images are used to select the ROI. This involves identifying and isolating areas likely to contain abnormalities, such as masses or calcifications. Techniques like morphological operations and edge detection accurately delineate these regions.

Contrast Enhancement Figure 6 illustrates the image after applying Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE is a powerful contrast enhancement technique based on Adaptive Histogram Equalization (AHE). It operates by computing the histogram for the local region surrounding each pixel, and then mapping the pixel intensity to a new value within the display range. This new value is proportional to the pixel intensity's rank in the local histogram, thereby enhancing the local contrast of the image.

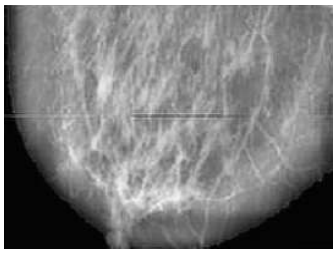


Figure 6. CLAHE filtered Image

Histogram of the image processed with the Wiener filter. This histogram represents the distribution of pixel intensities after noise reduction. The Wiener filter effectively minimizes noise while preserving significant image details, as reflected in the enhanced clarity and contrast of the histogram in Figure 7.

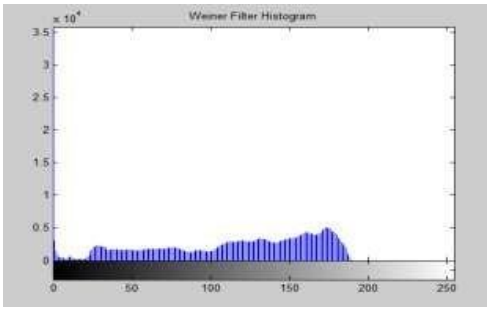


Figure. Weiner filter histogram

Histogram of the image processed with Contrast Limited Adaptive Histogram Equalization (CLAHE). This histogram demonstrates the distribution of pixel intensities following CLAHE application, highlighting the enhanced contrast and improved visibility of details in the image. The CLAHE technique optimizes local contrast, making subtle features more distinguishable.

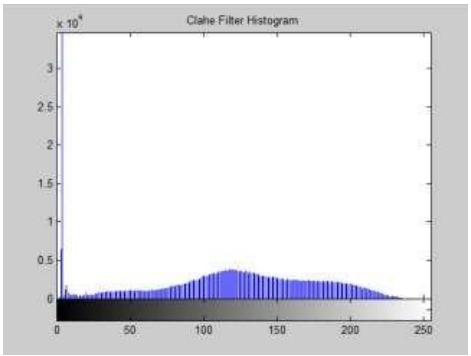


Figure 8: CLAHE filtered Histogram Image

Table 1 presents various statistical measures of an image processed using the CLAHE (Contrast Limited Adaptive Histogram Equalization) filter. The parameters include standard deviation, variance, mean, and maximum pixel value, providing a comprehensive overview of the image's characteristics after applying the CLAHE filter.

Table 1. Different values of CLAHE filter image

Sr.No.	Parameters	Values
1	STANDARD DEVIATION	1.366×10^04
2	Variance	1.866 ×10^08
3	Mean	2604
4	Maxvalue	219557

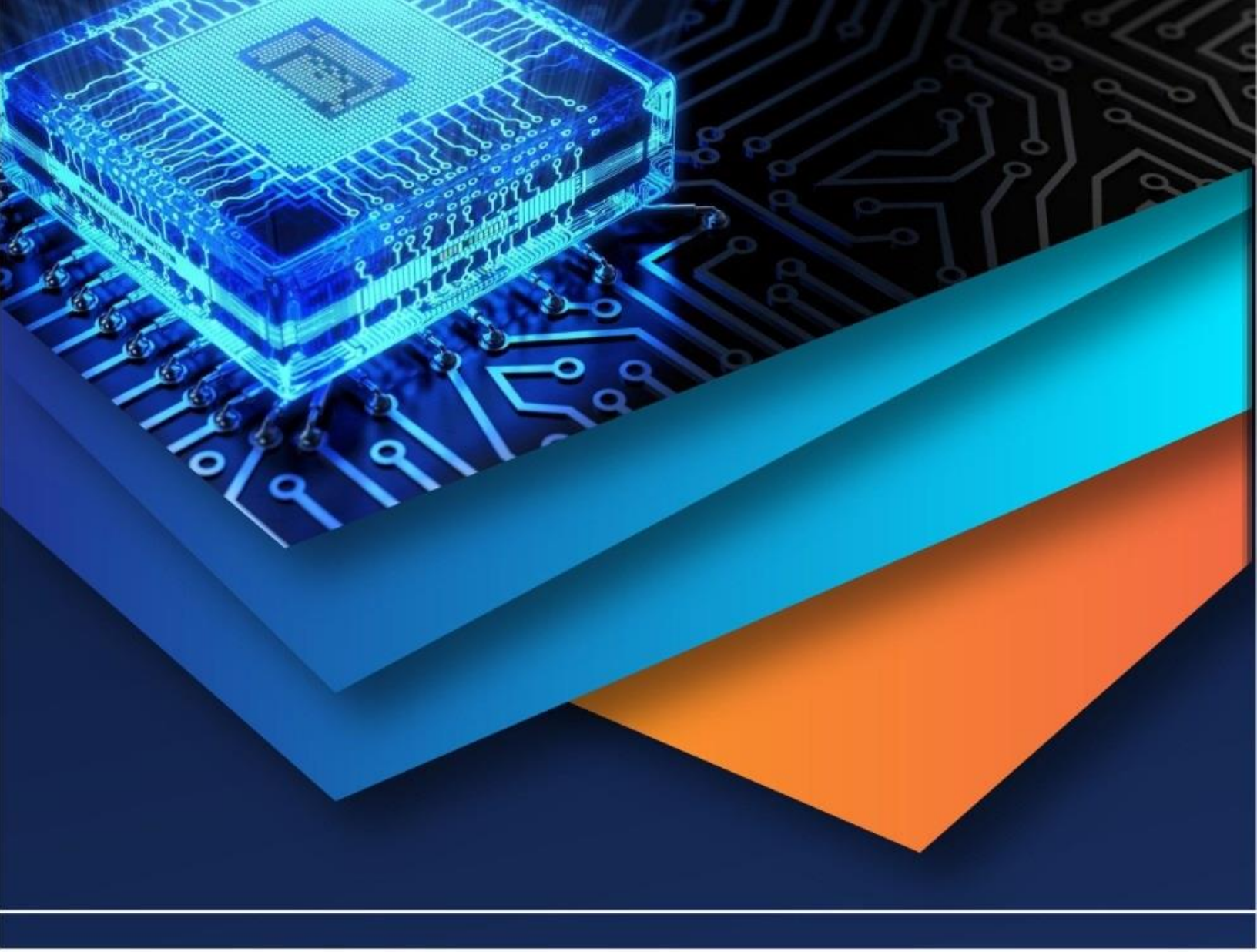
V. CONCLUSIONS

This investigation focused on detecting and calculating features in mammograms using Convolutional Neural Networks (CNNs) to distinguish between normal and abnormal mammograms. The study applied deep learning techniques on the Mammographic Image Analysis Society (MIAS) dataset, addressing feature extraction by refining the classification of abnormal cases to align with normal ones. Various filter sizes and preprocessing methods were employed to reduce noise and enhance the overall accuracy of the system. Effective segmentation proved essential for accurate feature extraction and classification. The use of morphological operations for segmentation notably improved the classification results. Consequently, the implementation demonstrated that images could be reliably analyzed and classified for further evaluation.

REFERENCES

[1] Lei Fana Paul E. Gossb, c Kathrin Strasser-Weippld, “Current Status and Future Projections of Breast Cancer in Asia”, Breast Care 2015;10:372–378.DOI: 10.1159/000441818
[2] https://www.repository.cam.ac.uk/handle/1810/250394
[3] “Breast cancer statistics.” [Online]. Available: http://www.wcrf.org/int/ cancer-facts-figures/data-specific-cancers/breast-cancer- statistics.A. Jemal et al., “Cancer statistics, 2008,” CA. Cancer J. Clin., vol. 58, no. 2, pp. 71–96, Apr. 2008.
[4] A. Paul and D. P. Mukherjee, “Mitosis detection for invasive breast cancer grading in histopathological images,” IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4041–4054, 2015.
[5] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, “Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1313–1321, 2016.
[6] B. E. Bejnordi, M. Veta, P. J. van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, J. A. van der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol et al., “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” Jama, vol. 318, no. 22, pp. 2199–2210, 2017.
[7] Y. Tan, K. Sim, and F. Ting, “Breast cancer detection using convolutional neural networks for mammogram imaging system,” in Robotics, Automation and Sciences (ICORAS), 2017 International Conference on. IEEE, 2017, pp. 1–5.
[8] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, “The digital database for screening mammography,” Digital mammography, pp. 431–434, 2000.
[9] N. Dhungel, G. Carneiro, and A. P. Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on. IEEE, 2015, pp. 1– 8.
[10] G. Carneiro, J. Nascimento, and A. P. Bradley, “Automated analysis of unregistered multi-view mammograms with deep learning,” IEEE transactions on medical imaging, vol. 36, no. 11, pp. 2355–2365, 2017.
[11] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, “Inbreast: toward a full-field digital mammographic database,” Academic radiology, vol. 19, no. 2, pp. 236–248, 2012.
[12] E. Castro, J. S. Cardoso, and J. C. Pereira, “Elastic deformations for data augmentation in breast cancer mass detection,” in Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on. IEEE, 2018, pp. 230–234.
[13] E. Kozegar, M. Soryani, H. Behnam, M. Salamati, and T. Tan, “Mass segmentation in automated 3-d breast ultrasound using adaptive region growing and supervised edge-based deformable model,” IEEE Transactions on Medical Imaging, 2017.
[14] S. Beura, B. Majhi, R. Dash, and S. Roy, “Classification of mammogram using two-dimensional discrete orthonormal s-transform for breast cancer detection,” Healthcare technology letters, vol. 2, no. 3, pp. 46–51, 2015.
[15] K. S. Beevi, M. S. Nair, and G. Bindu, “A multi-classifier system for automatic mitosis detection in breast histopathology images using deep belief networks,” IEEE journal of translational engineering in health and medicine, vol. 5, pp. 1–11, 2017.
[16] K. Hu, X. Gao, and F. Li, “Detection of suspicious lesions by adaptive thresholding based on multiresolution analysis in mammograms,” IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 2, pp. 462–472, 2011.
[17] M. Al-Ayyoub, S. M. AlZu’bi, Y. Jararweh, and M. A. Alsmirat, “A gpubased breast cancer detection system using single pass fuzzy c-means clustering algorithm,” in Multimedia Computing and Systems (ICMCS), 2016 5th International Conference on. IEEE, 2016, pp. 650–654

[18] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," IEEE transactions on medical imaging, vol. 35, no. 2, pp. 550–566, 2016.
[19] J. Shell and W. D. Gregory, "Efficient cancer detection using multiple neural networks," IEEE journal of translational engineering in health and medicine, vol. 5, pp. 1–7, 2017.
[20] H. Lin, H. Chen, Q. Dou, L. Wang, J. Qin, and P.-A. Heng, "Scannet: A fast and dense scanning framework for metastastic breast cancer detection from whole-slide image," in Applications of Computer Vision (WACV), 2018 IEEE Winter Conference on. IEEE, 2018, pp. 539–546.
[21] A. Elmoufidi, K. El Fahssi, S. Jai-andaloussi, A. Sekkaki, Q. Gwenole, and M. Lamard, "Anomaly classification in digital mammography based on multiple-instance learning," IET Image Processing, 2017.
[22] K. Hu, W. Yang, and X. Gao, "Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden markov tree model of dual-tree complex wavelet transform," Expert Systems with Applications, vol. 86, pp. 135–144, 2017.
[23] M. Al-Masni, M. Al-Antari, J. Park, G. Gi, T. Kim, P. Rivera, E. Valarezo, S.-M. Han, and T.-S. Kim, "Detection and classification of the breast abnormalities in digital mammograms via regional convolutional neural network," in Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE. IEEE, 2017, pp. 1230–1233.
[24] Souza, M.D., Prabhu, G.A., Kumara, V. *et al.* EarlyNet: a novel transfer learning approach with VGG11 and EfficientNet for early-stage breast cancer detection. *Int J Syst Assur Eng Manag* (2024). https://doi.org/10.1007/s13198-024-02408-6
[25] Melwin D'souza, Ananth Prabhu Gurpur, Varuna Kumara, "SANAS-Net: spatial attention neural architecture search for breast cancer detection", IAES International Journal of Artificial Intelligence (IJ-AI), Vol. 13, No. 3, September 2024, pp. 3339-3349, ISSN: 2252-8938, DOI:http://doi.org/10.11591/ijai.v13.i3.pp3339-3349
[26] Melwin D Souza, Ananth Prabhu G and Varuna Kumara, A Comprehensive Review on Advances in Deep Learning and Machine Learning for Early Breast Cancer Detection, International Journal of Advanced Research in Engineering and Technology (IJARET), 10 (5), 2019, pp 350-359. https://iaeme.com/Home/issue/IJARET?Volume=10&Issue=5



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)