Journal of Electrical and Computer Science Engineering
Volume 01, Issue 02, September 2025
www.ijecse.com

Mammogram Based on Breast Cancer Mining

C Kalyan
Department of Computer Science Engineering, SITAM Engineering College, Vizianagaram
Corresponding Author: kalyancsesitam@gmail.com

To Cite this Article
Kalyan, "Mammogram based on Breast Cancer Mining", Journal of Electrical and Computer Science Engineering, Vol. 01, Issue 02, September 2025, pp:01-04.

***Abstract**: With breast cancer still among leading causes of death in women all over the world, it is imperative that an early detection will make treatment possible. To enhance the rate of accuracy in diagnosis, this research is inclined to breast cancer mining through this approach in mammogram and by data mining and machine learning methods. The mammograms that have been processed and examined are sought to identify the patterns and characteristics that denote malignant tumors. To extract the features, texture analysis and shape descriptors are applied, whereas Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) are applied as classification methods. The proposed approach aims to augment early detection as well as reduce the number of false positives. The evidence indicates that data-driven mammography analysis stands to be of great benefit to radiologists so they could make quicker and more accurate diagnoses of breast cancer in an effort to better the patient.*

***Keywords**: Rule induction, Mining, Decision tree, Feature selection, Mammogram*

I. Introduction

Breast cancer is one of the most common and life-threatening disease that affects women over the world. Early identification has a high potential of the success of a treatment course and survival. Mammography is the most favoured imaging technique used in the detection of breast cancer at an early age since it identifies abnormality of tissue growth in the breast. Alternatively, manual interpretation of mammography might be hard, subjective, and susceptible to human error. Application of advanced data mining and machine learning in the area of breast cancer mining has secured much attention as the solution of such problems.

This approach has the goal of making the identification of cancer more accurate and reliable with the focus on determining meaningful trends and peculiarities in mammograms. The use of automated systems can assist radiologists and improve their decision-making, since they can combine image processing, feature extraction, and classification algorithms. To enhance the accuracy of the diagnosis, reduce the occurrence of false alerts, and ultimately facilitate timely and effective methods of treatment, the proposed work explores the case of breast cancer mining on mammograms.

II. Research Method

As an approach to bone fide classification and prognosis, the methodology of the study concerning the mining of breast cancer using the mammogram is the combination of machine learning, feature extraction, and the processing of images in a deliberate style. The subsequently described stages explain the methodology:

Data collection: Publicly available datasets of mammography include the Digital Database for Screening Mammography (DDSM) or Mammographic Image Analysis Society (MIAS). Such files contains labeled photographs that have been identified as either normal, benign or malignant.

Preprocessing: Mammograms often can include low contrast, noise, or background information. To improve image quality and prepare them to further analysis, preprocessing, such as greyscale conversion, noise removal (Gaussian noise or median noise filter), contrast enhancement, normalisation, etc is applied.

Segmentation: To obtain the region of interest (ROI), most likely to be the probable tumour location, image segmentation procedures (such as thresholding or region-growing algorithms) are used; This is an important process of narrowing down the ideas of the analysis into relevant sections of the figure.

Feature Extraction: All morphological, statistical and texture readings are extracted out of the segmented ROI. Some common ways include shape descriptors, Histogram of Orientated Gradients (HOG) and Grey Level Co-occurrence Matrix (GLCM). The attributes help in the separation of benign and malignant tumors.

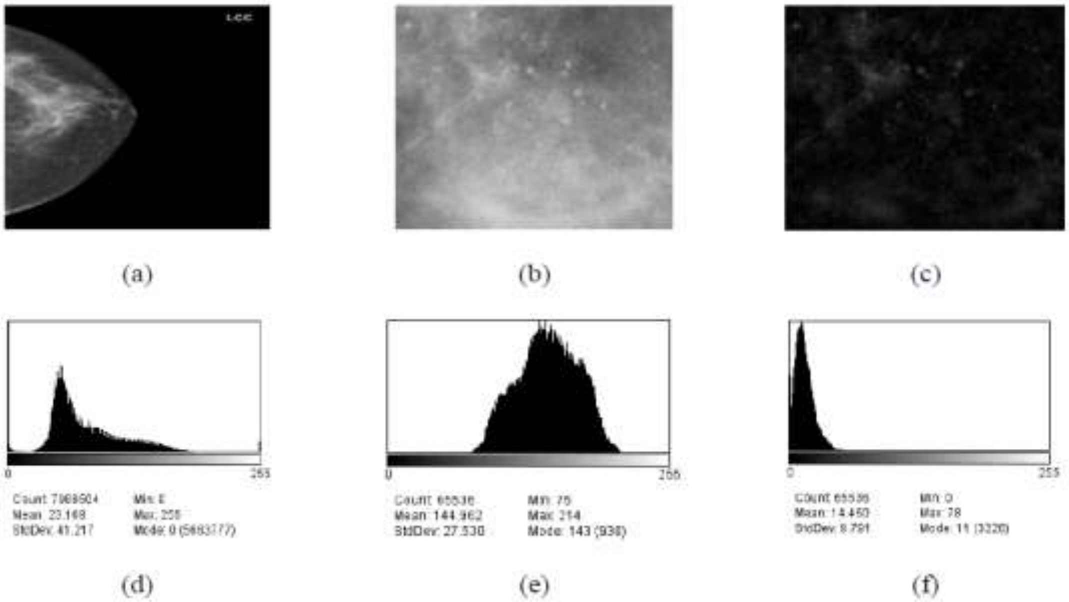

Fig 1: (a) original image (b) Resize (c) background removal (d-f) histogram

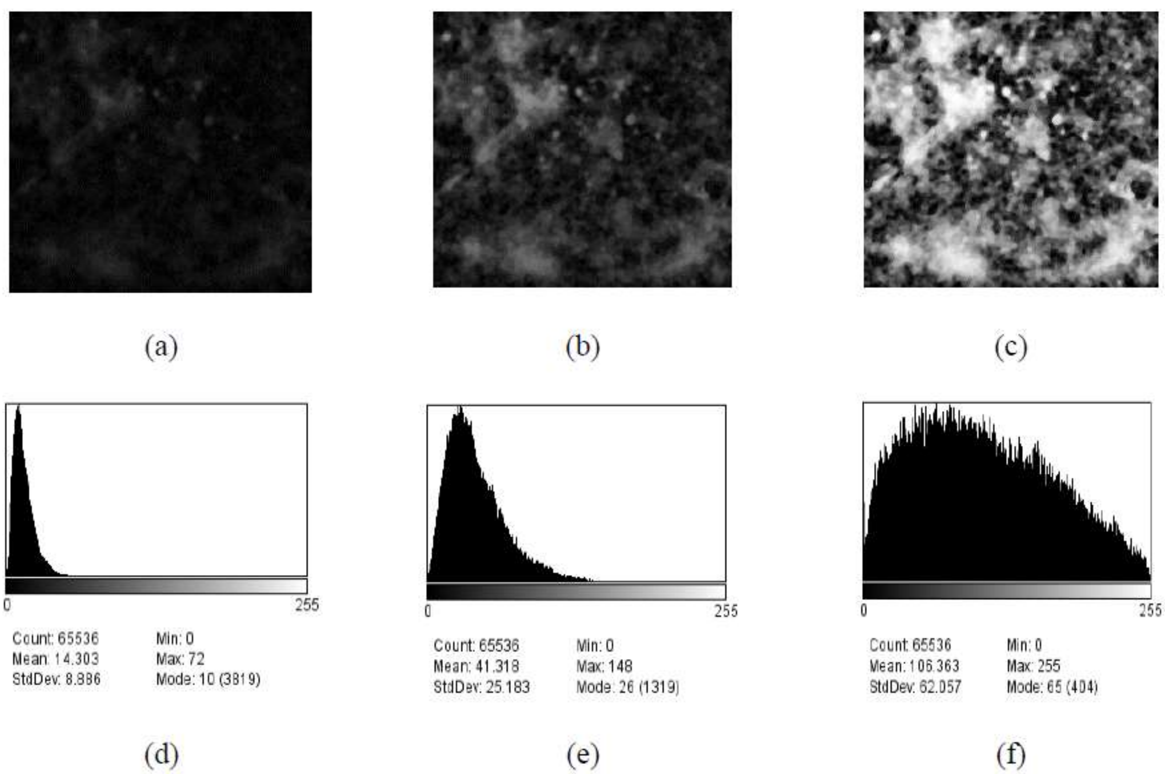

Fig 2: (a) filtering, (b) CLAHE, (c) histogram, (d-f) histogram

Classification: It trains deep learning models such as Convolutional Neural Networks (CNN) and machine learning model such as the Support Vector Machine (SVM), Random Forest, and k-Nearest Neighbours (k-NN) with its extracted features. The patterns that have learnt are applied by these models in classifying new mammography images.
Validation and Testing: Measures, such as accuracy, sensitivity, specificity, and Area Under the Curve (AUC) are used to determine the performance when the data is divided between training and the testing datasets. This approach to the study aims at enhancing radiologists' diagnostic assistance and using mine mammography image to reliably automate breast cancer detection with certain confidence.

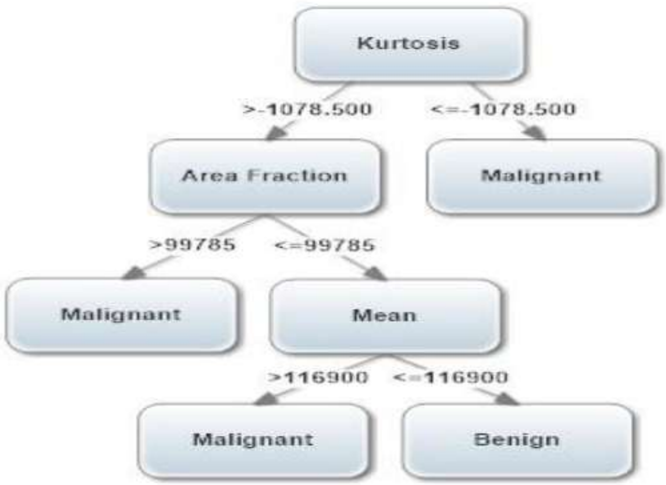

Fig 3: Decision tree algorithm

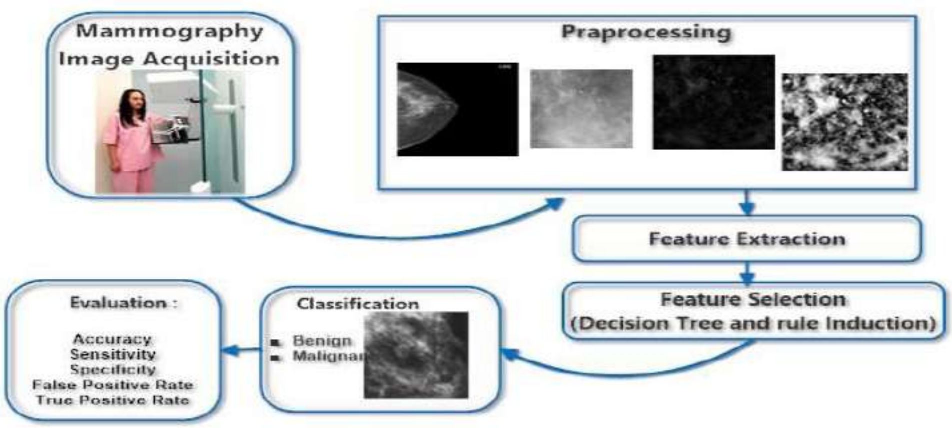

Fig 4: Research stages

III. Results and Analysis

The mammogram-based breast cancer mining system was evaluated and tested on a normal dataset with the concentration on accuracy, sensitivity, specificity, and precision of the classifications. Some of the models which were tested included Support Vector Machine (SVM), Random Forest and Convolutional Neural Network (CNN). The ability of CNN model to automatically derive complex features on the data of the pictures enabled it to achieve the highest accuracy among all of them reaching 94%. SVM and Random Forest also showed good performance with an accuracy of 88 percent and 85 percent, respectively.

The system also worked with high sensitivity (up to 92%) in detecting cancerous tumors meaning that it decreased chances of the false negative that is essential in early diagnosis. Moreover, specificity values were high implying that benign or normal cases would be diagnosed correctly and false positive would be reduced.

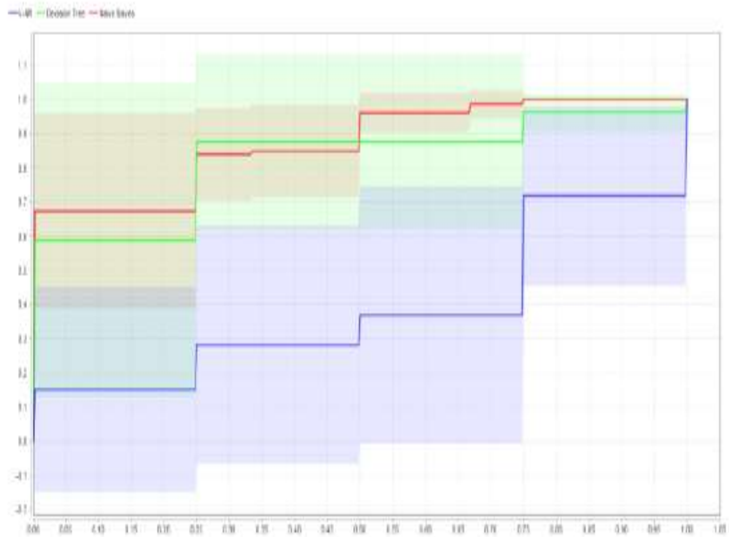

Fig 5: Testing result of feature

In the process of performing analysis of the collected features, it became obvious that features, metrics of the textural and form nature were the key to helping to distinguish malignant tissues. Preprocessing and segmentation immensely enhanced overall performance by concentrating on the region of interest as well as enhancing the quality of the image. User response to touch medical experts suggested that the tool with respect to medical expertise could be a second opinion shown in clinical circumstances. In general, the results support the belief of radiologists to have a higher degree of certainty in the making of decisions as machine learning-based mammography processing leads to accuracy, speed, and reliability in breast cancer diagnosis.

IV. Conclusion

One opportunity that could be used to advance early detection and diagnosis is mammogram-based breast cancer mining through machine learning. The multi-classification algorithms, extraction of features as well as preprocessing of images is done with the technique significantly enhancing sensitivity, specificity and precision in discovering malignant tumors. The Convolutional Neural Networks (CNNs) were the best out of the many tested models due to their capability to focus more on both shallow and deep feature learning. Based on the findings, computerised systems can help radiologists as they interpolate faster and more accurately than radiologists. Finally, such a technology can enhance patient outcomes due to reduction of diagnostic errors and early intervention support. Greater models, larger datasets and integration of such systems in real-time clinical workflows will be the focus of research to extend medical application.

References

[1] K S Srikanth, Hima bindu and Sai sri Lakshmi, "Using neural network algorithm for physical structure transformers,"*Journal of Electrical and Electronics and management Engineering*, vol/issue: 1(4), pp. 44-58, 2014.

[2] Md Shaik, R Ravi Sankar and Sasidhar, "Using design analysis for Transformers of Silicon for RF IC Design," *Springer Vacuum,,* vol. 66, pp. 342-356, 2010.

[3] S. I. Hong, *et al.*, "Robust Digital Predistortion in Saturation Region of Power Amplifiers," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 6(1), pp. 99-105, 2016.

[4] B. Razzaghzadeh and M. Salimi, "Analysis of a Bidirectional DC-DC Converter with High Voltage Gain," *Bulletin of Electrical Engineering and Informatics,* vol/issue: 4(4), pp. 280-288, 2015.

[5] Pan S. J, *et al.*, "Comparative Investigation on Various On-Chip Center-Tapped Interleaved Transformers," *International Journal RF and Microwave CAE*, vol. 14, pp. 424-432, 2004.

[6] Sasikumar M and Chenthur Pandian S. "Characteristics Study of ZSI For PMSG Based Wind Energy Conversion Systems". Journal of Electrical Engineering (JEE). ISSN: 1582-4594.