IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

BREAST CANCER DETECTION USING MACHINE LEARNING TECHNIQUES

Pradeeba.R,

Msc Computer science,

Nirmala College for women, Coimbatore.

Mrs.C.Clement Sherlin, MCA, MPhil,

Assistant Professor, Department of Computer science,

Nirmala College for women, Coimbatore.

Abstract: Breast cancer is one of the leading causes of death in women. Many studies on the diagnosis and detection of breast cancer have been conducted using various image processing and classification techniques. Despite this, the disease remains one of the deadliest. Having given birth to one out of every six women in her lifetime. Early detection of breast cancer tumours is the only way to cure it. The most efficient and simple way to diagnose breast cancer is to use CAD (Computer Aided Diagnosis) on mammographic images. Accurate discovery can significantly reduce the mortality rate caused by mamma cancer. Clusters of masses and microcalcifications are important early symptoms of breast cancer. They can aid in the early detection of breast cancer. This paper quantifies the image analysis methods used for tumour detection. These tumours are extracted from the mammogram ROI using FCM to classify the microcalcifications as harmless, ordinary, or dangerous. These images are compared further and run through CNN to gain a better understanding of the cancer pattern in the mammography image.

Keywords: Breast cancer, CAD, CNN, FCM, Mammographic image

I. INTRODUCTION

Cancer is the name given to a group of abnormal cells that begin to divide and grow uncontrollably, spreading into the surrounding tissues and forming tumours [1]. These tumours can develop in any organ of the human body, and anyone can develop breast cancer. Breast cancer is the most common cancer in women all over the world. As a result, early detection and treatment of cancer reduces the risk of death while increasing survival rates. Mammography, ultrasound, Magnetic Resonance Imaging (MRI), and thermography are all imaging techniques used to detect breast cancer. Mammograms are less effective and have some drawbacks, such as being painful, invasive, and ineffective in the case of women with dense breasts. MRI, like mammography, cannot distinguish between a cancerous lump and a benign cyst. Any Computer-Aided Detection (CAD) system for breast cancer detection can be divided into three stages: pre-processing and ROI, feature extraction, classification, and performance analysis. The goal of ROI segmentation is to separate the breast region from the rest of the thermal images. Following this segmentation, some features are extracted. Some artificial classifications algorithms are typically used to classify the breast as normal or abnormal [2]. Typically, the accuracy of the segmentation result determines the classification results of breast thermal images. The accurate segmentation of regions of interest from medical thermal images is still a work in progress. Thermal images have inherent limitations such as the lack of clear edges, low contrast, and a low signal to noise ratio. As a result of the more intricate intensity field of the breast thermogram, complex pre-segmentation steps remain a challenge. An automatic breast cancer detection technique is presented in this work. The proposed novel segmentation technique that uses the concave and convex shape of the frontal view of the breast is the main contribution of the presented work. To assess the accuracy of the proposed method, different selective features were extracted from the segmentation regions and a neural network (NN) classifier was used to detect breast abnormalities.

II. LITERATURE REVIEW

This article [3] provides a comparison of machine learning, deep learning, and data mining techniques used for breast cancer prediction. Our primary goal is to compare various existing Machine Learning and Data Mining techniques in order to identify the most appropriate method for supporting large datasets with high prediction accuracy.

We identify breast cancer-related genes with significantly better performance than other existing machine learning methods in this paper [4]. The predicted genes with prognostic values may play important roles in breast cancer and may be candidates for future biomarker studies by biologists and medical scientists.

This study [5] optimises the survival analysis of the XGBoost framework for ties to predict breast cancer disease progression. Methods: EXSA is built on the XGBoost machine learning framework and the Cox proportional hazards model for survival analysis. The authors created the EXSA method to create an excellent prognostic model for estimating disease progression.

The authors of this paper [6] address the problem of breast cancer prediction in the context of big data. The authors looked at two types of data: gene expression (GE) and DNA methylation (DM). The goal of this paper is to scale up classification machine-learning algorithms by applying each dataset separately and jointly. The experimental results demonstrated that the scaled SVM classifier in the Spark environment outperforms the other classifiers, achieving the highest accuracy and lowest error rate with the GE dataset.

Breast cancer is a particularly aggressive form of cancer with a very low median survival rate. Accurate breast cancer prognosis prediction can save a significant number of patients from receiving unnecessary adjuvant systemic treatment and its associated high medical costs. Previous research has mostly relied on selected gene expression data to build a predictive model. The emergence of deep learning methods and multi-dimensional data allows for more comprehensive analysis of the molecular characteristics of breast cancer, potentially improving diagnosis, treatment, and prevention. In this study, we propose a Multimodal Deep Neural Network (MDNNMD) for breast cancer prognosis prediction by integrating Multi-dimensional Data.

III. METHODOLOGY

Proposed System

Mammography is the primary screening and early diagnosis tool, and proper interpretation of the clinical report is critical for breast cancer prediction, but the decision is subject to error. In this paper, we propose a method for efficiently examining digital mammograms using FCM segmentation to detect early-stage tumours that are not susceptible to human error, allowing us to take appropriate action to reduce the risks associated with breast cancer.

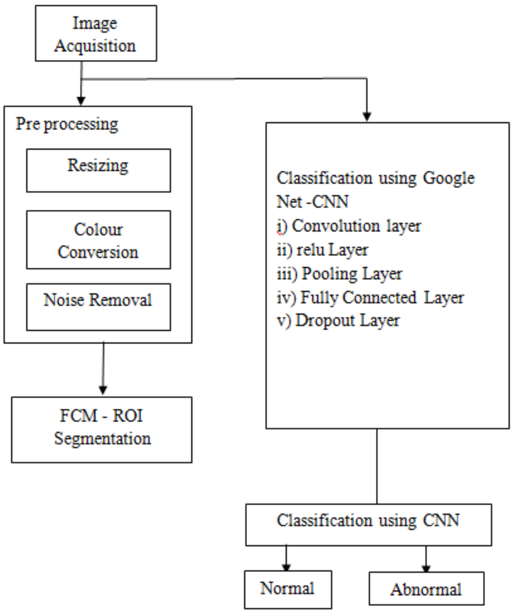


Fig. 1. Proposed Architecture

Image Acquisition

Image Acquisition is the process of obtaining an input image for the automatic detection of breast cancer via Digital Image Processing.

Pre-processing

At the most basic level of abstraction, both input and output are intensity images, which are referred to as pre-processing. The goal of pre-processing is to improve image data by suppressing unwanted distortions or enhancing some image features that are important for subsequent processing..

Noise Removal using Median Filter

The median filter is a nonlinear digital filter used to remove noise from an image or signal. A common pre-processing step used to improve the results of subsequent processing is noise reduction (for example, edge detection on an image).

Segmentation

Image segmentation is commonly used to find objects and boundaries (lines, curves, and so on) in images. Image segmentation is the process of labelling every pixel in an image so that pixels with the same label share certain characteristics..

FCM

It is used to segment an image by grouping pixels with similar or nearly similar values into clusters, where each group of pixel values in one cluster is similar to each other but different from pixel values in other clusters, and then these clusters represent the segments of the segmented image.

Classification

Image classification examines the numerical properties of various image features and categorises data. Classification algorithms are typically processed in two stages: training and testing. During the initial training phase, typical image feature characteristics are isolated and used to create a unique description of each classification category, i.e. training class. These feature-space partitions are then used to classify image features in the subsequent testing phase.

Working of CNN model for Breast cancer detection

➢ Layer of CNN model:

- ❖ Convolution 2D
- ❖ MAX Poolig2D
- ❖ Dropout
- ❖ Flatten
- ❖ Dense
- ❖ Activation

➢ Convolution 2D: In the Convolution 2D extract the featured from input image. It given the output in matrix form.

➢ MAX Poolig2D: In the MAX polling 2D it takes the largest element from rectified feature map.

➢ Dropout: Dropout is randomly selected neurons are ignored during training.

➢ Flatten: Flatten feed output into fully connected layer. It gives data in list form.

➢ Dense: A Linear operation in which every input is connected to every output by weight. It followed by nonlinear activation function.

➢ Activation: It used sigmoid function and predicts the probability 0 and 1.

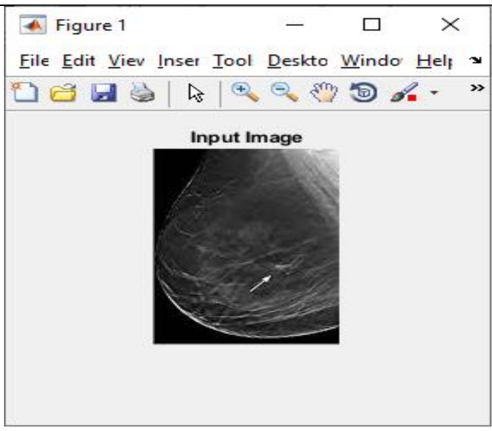
➢in the compile model we used binary cross entropy because we have two layers 0 and 1.

➢We used Adam optimizer in compile model. Adam:-Adaptive moment estimation. It used for non convex optimization problem like straight forward to implement.

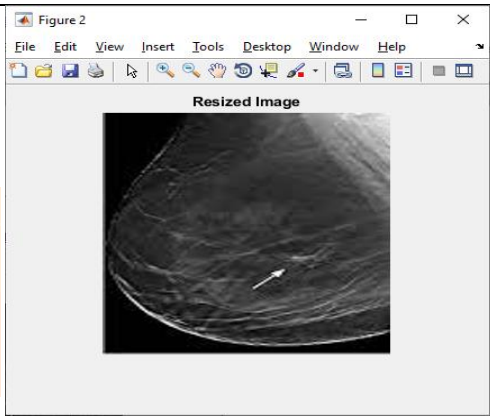
- Computationally efficient.
- Little memory requirement.

IV RESULTS AND DISCUSSION

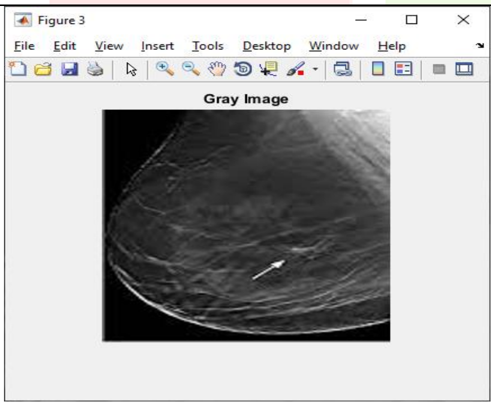
The following figures shows the results obtained from this work.



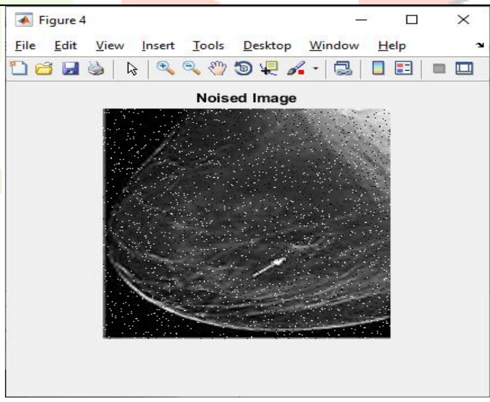
(a)



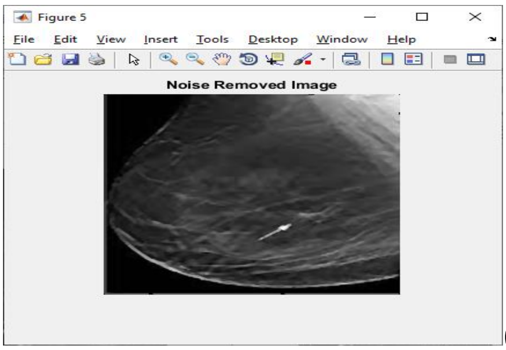
(b)



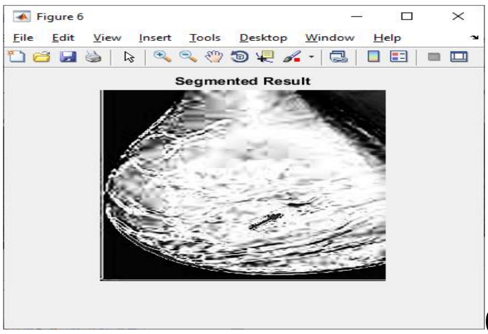
(c)



(d)



(e)



(f)

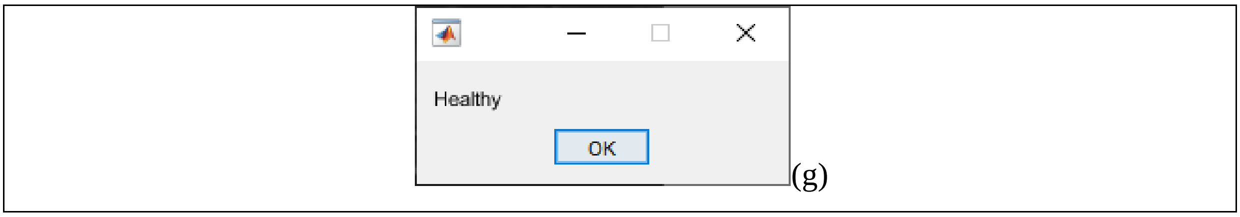
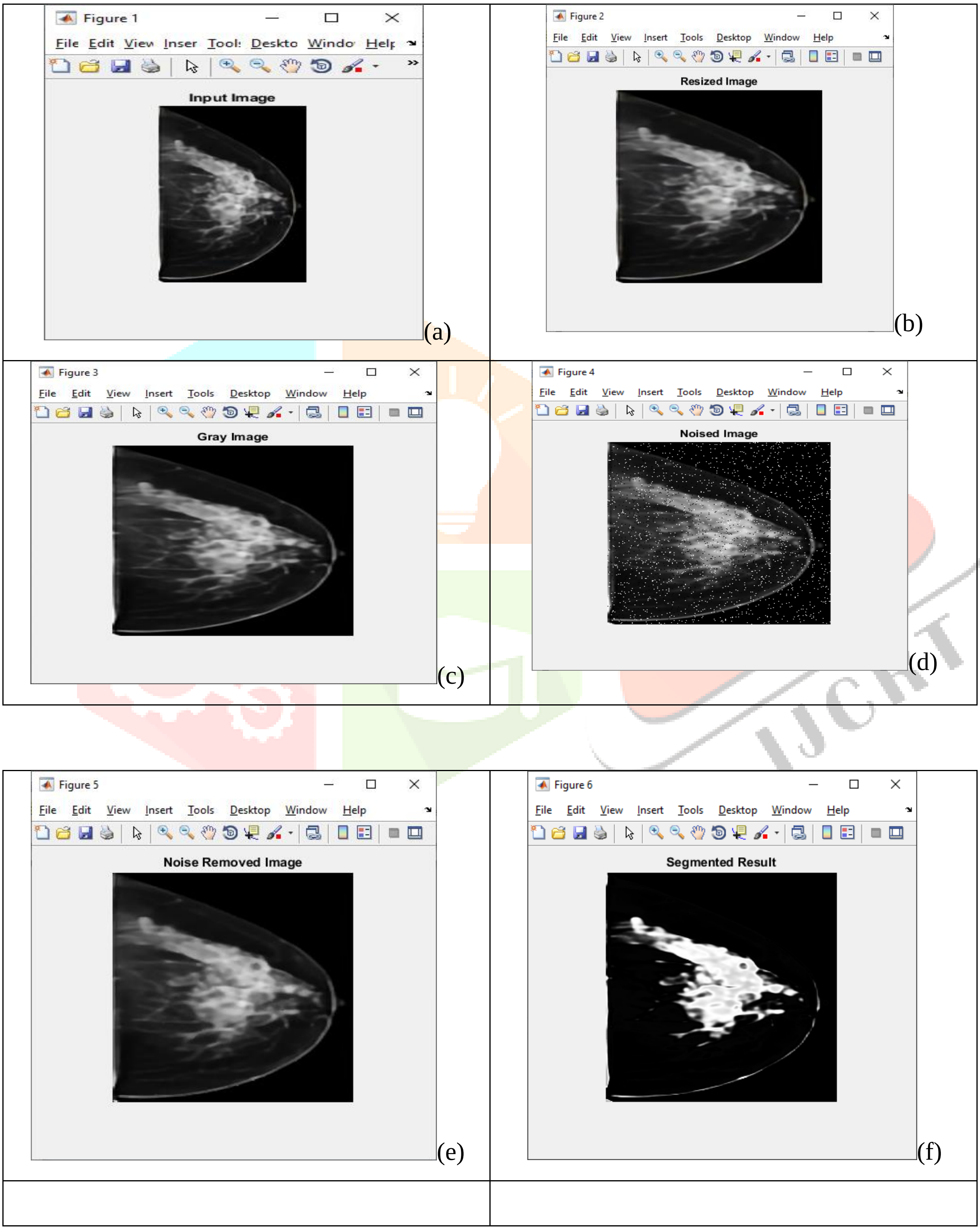
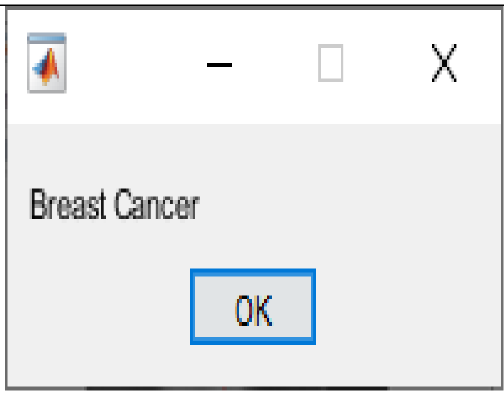


Fig.2. (a) to (g) Steps involved in predicting the breast cancer. (a) Input image (b) Resized image (c) Gray image (d) Noised image (e) Noise removed image (f) Segmented result (g) Result



(g)



(f)

Fig.3. Fig.2. From (a) to (g) Steps involved in predicting breast cancer. (a) Input image (b) Resized image (c) Gray image (d) Noised image (e) Noise removed image (f) Segmented image (g) Showing diseased part (f) Result

CONCLUSION & FUTURE WORK

This study compares various machine learning algorithms for detecting breast cancer from a digitised image of a fine needle aspirate (FNA) of a breast mass. The simple, safe, accurate, and inexpensive FNA procedure combined with the predictive model in this paper can be used for prognosis, diagnosis, and assisting doctors in making more accurate final decisions in less time with less human and monetary resources. The deep learning models' performance shows promising results for a near-perfect detection system. However, the lack of data instances and the conversion of data in order to use it for CNN proved to be a difficult part of the process..

The model can be improved in the future as data availability and, more importantly, data growth increase.

REFERENCES

[1] J. Ferlay, C. Héry, P. Autier and R. Sankaranarayanan, Global Burden of Breast Cancer, New York, NY, USA:Springer, pp. 1-19, 2010.

[2] L. R. Martin, S. L. Williams, K. B. Haskard and M. R. DiMatteo, "The challenge of patient adherence", Ther. Clin. Risk Manage., vol. 1, no. 3, pp. 189-199, 2005.

[3] N. Fatima, L. Liu, S. Hong and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," in IEEE Access, vol. 8, pp. 150360-150376, 2020, doi: 10.1109/ACCESS.2020.3016715.

[4] Peng, C., Zheng, Y., & Huang, D. S. (2020). Capsule Network Based Modeling of Multi-omics Data for Discovery of Breast Cancer-Related Genes. IEEE/ACM transactions on computational biology and bioinformatics, 17(5), 1605–1612. https://doi.org/10.1109/TCBB.2019.2909905

[5] Liu, P., Fu, B., Yang, S. X., Deng, L., Zhong, X., & Zheng, H. (2021). Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer. IEEE transactions on bio-medical engineering, 68(1), 148–160. https://doi.org/10.1109/TBME.2020.2993278.

[6] 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, doi: 10.1109/ACCESS.2019.2927080.

[7] Sun, D., Wang, M., & Li, A. (2018). A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM transactions on computational biology and bioinformatics, 10.1109/TCBB.2018.2806438. Advance online publication. https://doi.org/10.1109/TCBB.2018.2806438