medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Neural network pattern recognition of ultrasound image gray scale intensity histogram of breast lesions to differentiate between benign and malignant lesions

Arivan Ramachandran[1], Shiva Balan KR[2], Swathi Kiran[3], Mohamed Azharudeen[4]

1- Junior resident, Department of Nuclear Medicine, PGIMER, Chandigarh

2- Assistant professor, Department of General Medicine, MGMC, Puducherry

3- Junior resident, Department of Radiodiagnosis, JIPMER, Puducherry

4- Senior resident, Department of General Medicine, MGMC, Puducherry

ABSTRACT

The aim of this study is to analyze the effectiveness of grayscale intensity histogram to differentiate benign and malignant lesions using a convolutional neural network. Data (200 USG images, 100-malignant, 100-benign) was downloaded from an online access repository. The images were despeckled using ImageJ software and the grayscale intensity histogram values were extracted. In-built neural network pattern recognition application in Matlab R2019b was used to classify the images, which is a two-layer feed-forward network, with sigmoid hidden and softmax output neurons. The positive predictive value of the CNN was 95%. The best performance of 0.078264 was achieved at 36 epochs in the validation set. This study suggests that the grayscale intensity histogram of a USG image is an easy and feasible method to identify malignant lesions through an artificial neural network.

INTRODUCTION

Breast cancer is the most common cancer in Indian women with a prevalence of 25.8 per 1,00,000. Lack of adequate breast cancer screening, diagnosis at a later stage, and unavailability of resources are quoted as the main reasons for the increase in mortality in breast cancer patients in India (1). Multiple imaging

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

modalities like USG, X-ray mammography, CT, PET, and MRI are being used to screen, diagnose, and evaluate breast cancer.

Ultrasound is the imaging modality of choice in suspicious breast lesions in young women and in pregnant women. Ultrasound has higher accuracy and sensitivity in the detection of malignant lesions compared to X-ray mammography (2).

Usage of Artificial Intelligence (AI) for image recognition and classification is an upcoming method and can be implemented in areas with resource and man-power limitations as it is suggested that neural network-based differentiation of breast lesions has the capacity to significantly reduce unnecessary biopsies and can perform equivalent to trained human radiologists (3,4). In this study, we are evaluating the efficiency of Convolutional Neural Network (CNN) in classifying malignant and benign breast USG images downloaded from an online dataset based on their grayscale intensity histograms.

MATERIALS

- 200 USG Images (100 malignant and 100 benign) downloaded from https://data.mendeley.com/datasets/wmy84gzngw/1
- Matlab R2019b software
- ImageJ software

VARIABLE: Gray-scale intensity histogram

METHODOLOGY

Ultrasound images of 100 malignant and 100 benign breast lesions were randomly downloaded from an open access repository https://data.mendeley.com/datasets/wmy84gzngw/1(5). The images

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

were then loaded in the ImageJ software and despeckling was done to improve the contrast resolution of the images (6).

Region of interest (ROI) was drawn over the breast lesions in all the 200 images by a board-certified radiologist and the gray-scale intensity histogram values were extracted. The values were entered in a datasheet and were imported to the Matlab R2019b software.

A total of 200 histograms were divided into

1.Training set containing 70% of total images (140 out of 200)

2.Validation set containing 15% of total images (30 out of 200)

3.Test set containing 15% of total images (30 out of 200)

The in-built application of Matlab R2018b Neural net recognition was used. It is a two-layer feed-forward network, with sigmoid hidden and softmax output neurons. The network was trained with scaled conjugate backpropagation. In our study, we used 30 hidden neurons (7). An input file was created with 200 histogram values as a 255x200 matrix datasheet. The target output was assigned as 1 for malignant and 0 for benign using a 2x200 matrix datasheet. The software trained the CNN by a randomized selection of the inputs automatically.

RESULTS

The automated results were obtained, and the following figures were created.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

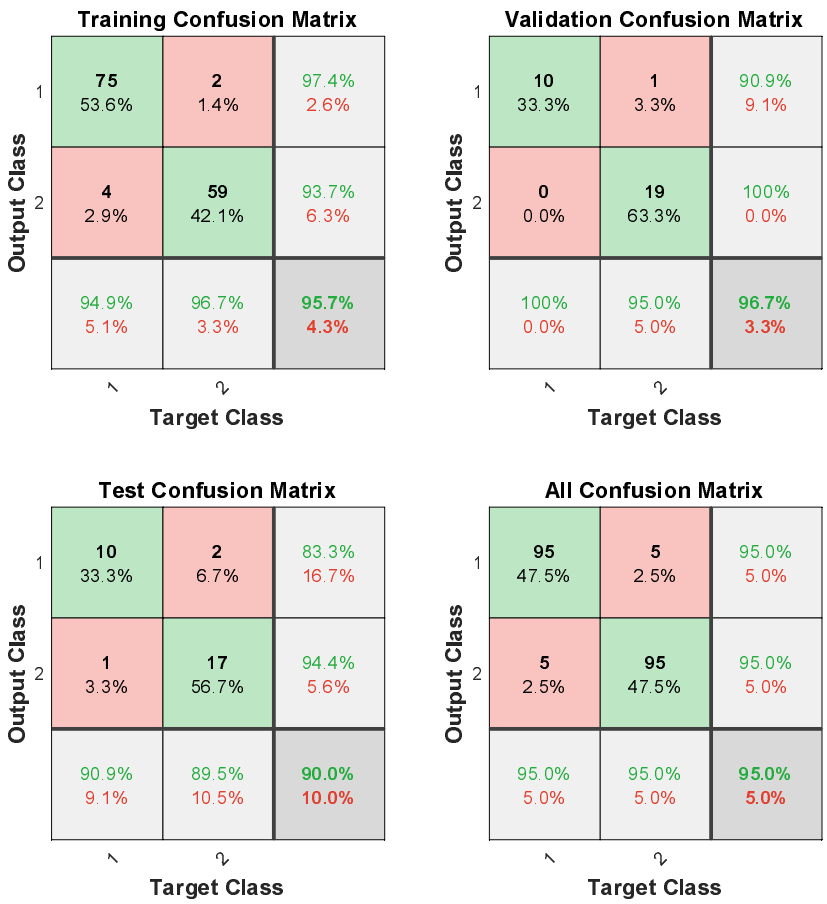

FIGURE.1- Confusion matrices of training, validation, test and combined sets

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

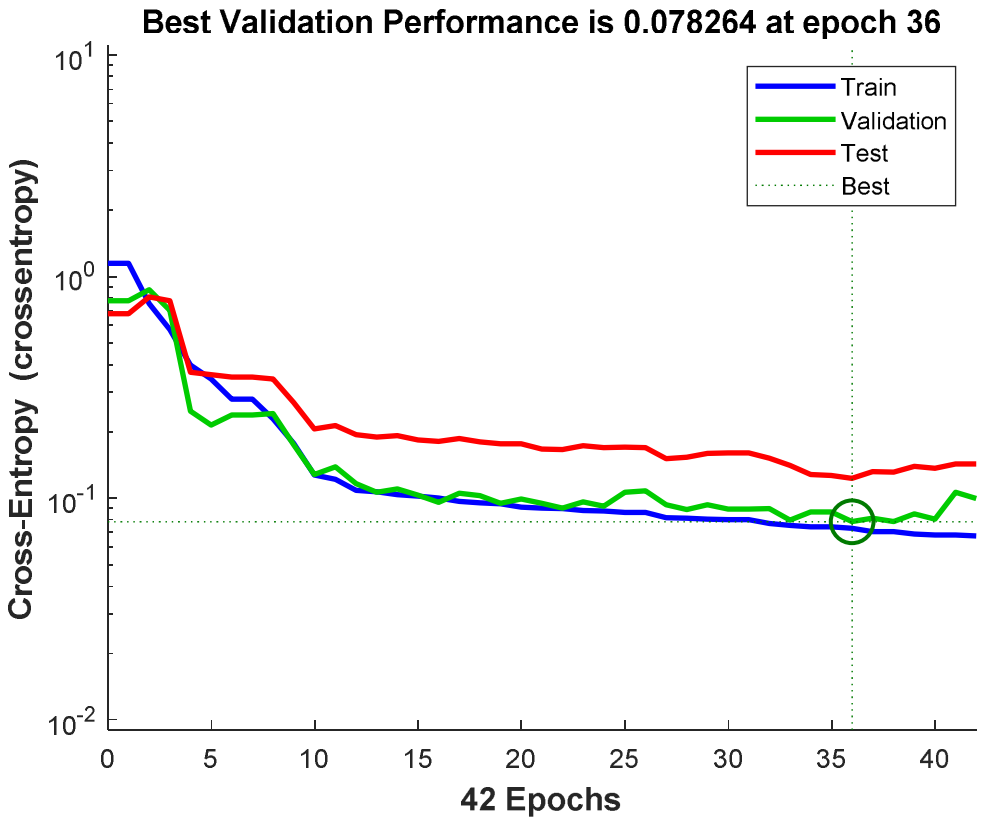

FIGURE.2- Performance plot of Matlab2019b CNN

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

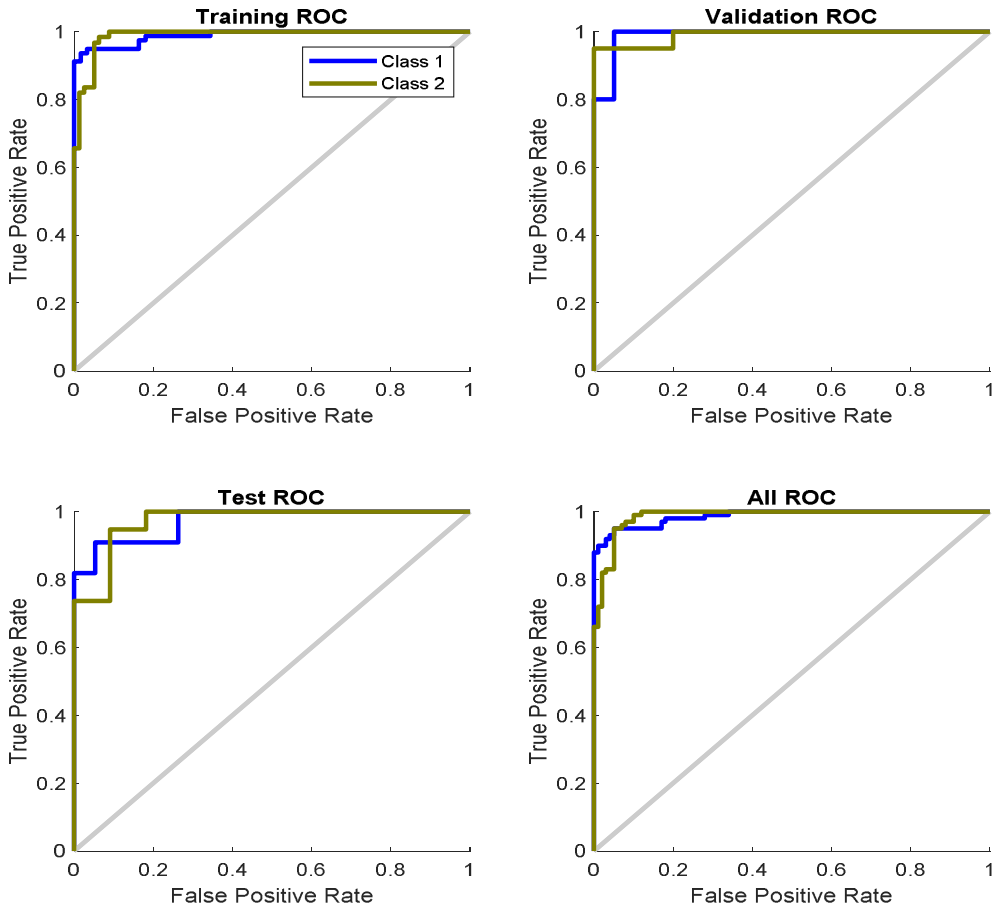

FIGURE.3- ROC (Receiver Operator Characteristic) curves

The prediction percentages were recorded in a confusion matrix. The positive prediction in the training set, validation set, test set and the combined set is 95.7%, 96.7%, 90% and 95% respectively (Figure.1).

The best performance of 0.078264 was achieved at 36 epochs in the validation set after which the cross-entropy increased, suggesting an increase in training errors (Figure.2)

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Based on the ROC curves (Figure.3), the positive predictive value was 95%.

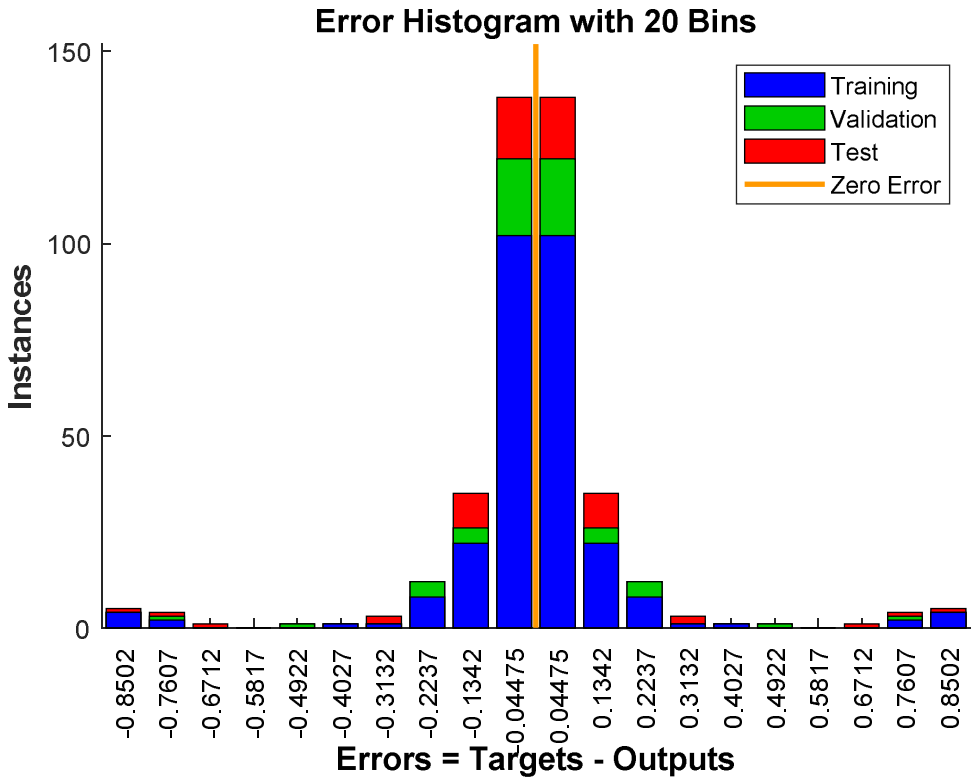

FIGURE.4 – Error Histogram

DISCUSSION

Artificial Intelligence aided analysis of images to differentiate lesions is an area of interest in the current era. The implementation of neural networks in differentiating malignant and benign breast lesions using grayscale intensity histogram is yet to be explored. In this study, we proved the significance of grayscale intensity histogram and its usage in neural networks to classify benign and malignant lesions.

We have analyzed only one ultrasound parameter of the lesions and the conclusions were derived. The ROIs were drawn manually in this study which is time-consuming and can result in inter-observer variability. The above-mentioned

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

limitations can be corrected by including multiple imaging variables that correlate with clinical findings and by using automated systems respectively.

The positive predictive value of the CNN was 95%. This suggests that the Convolutional Neural Network (CNN) analysis of the grayscale intensity histogram of a USG image of breast lesion is an easy and effective method to differentiate between malignant and benign nature of the lesion. Further studies are needed to increase the robustness of the findings of our study.

REFERENCES

1. Malvia S, Bagadi SA, Dubey US, Saxena S. Epidemiology of breast cancer in Indian women. Asia Pac J Clin Oncol. 2017 Aug;13(4):289-295.
2. Tan KP, Mohamad Azlan Z, Rumaisa MP, Siti Aisyah Murni MR, et al. The comparative accuracy of ultrasound and mammography in the detection of breast cancer. Med J Malaysia. 2014 Apr;69(2):79-85.
3. Goldberg V, Manduca A, Ewert DL, Gisvold JJ, Greenleaf JF. Improvement in specificity of ultrasonography for diagnosis of breast tumors by means of artificial intelligence. Medical physics. 19(6):1475-81, 1992.
4. Dumane VA, Shankar PM, Piccoli CW, Reid JM, Forsberg F, Goldberg BB. Computer aided classification of masses in ultrasonic mammography. Medical Physics. 29(9):1968-73, 2002.
5. Rodrigues, Paulo Sergio (2017), “Breast Ultrasound Image”, Mendeley Data, v1 http://dx.doi.org/10.17632/wmy84gzngw.1.
6. Loizou, Christos & Pattichis, et al. (2011). Despeckling. 10.1007/978-1-84882-688-5_6.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.01.20088245; this version posted May 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

7. MATLAB. 9.7.0.1190202 (R2019b). Natick, Massachusetts: The MathWorks Inc.; 2018.
8. Song, Jae & Venkatesh, Santosh & Conant, Emily & Cary, Ted & Arger, Peter & Sehgal, Chandra. (2005). Artificial Neural Network to aid differentiation of malignant and benign breast masses by ultrasound imaging. Proceedings of SPIE - The International Society for Optical Engineering. 5750. 10.1117/12.595295.