A CNN Based Breast Tumor Classifier Using Mendeley BUS Dataset

S.Sri Durga Kameswari, S.Sri Durga Kameswari, V Vijayakumar

Abstract: Women account to nearly fifty percent of the world population and most of them were being diagnosed with breast cancer. In the last decade, the worldwide breast cancer rate has been increased by more than twenty percent. A tumor may be cancerous or non-cancerous. There is a severe need to categorize the breast tumor for further treatment. Computer aided identification techniques can minimize the number of unnecessary biopsies. In this paper, a deep learning algorithm is proposed to label the tumors as benign or malignant. It uses three layers within it. We have implemented our model on breast ultrasound images. Mendeley dataset was used for this experiment which contains 250 ultrasound images of breast tumor out of which 100 were benign and 150 were of malignant class. The Proposed CNN was trained from scratch, trained with 50 epochs yields test accuracy of approximately 98%.

Index Terms: breast cancer, deep learning, ultrasound, CNN

I. INTRODUCTION

According to the statics provided by National Institutes of Cancer prevention and Research for the year 2018, Cancer is the second most common cause of death in India. Out of various types of cancers Breast cancer is most dominant type among Indian women and it has one-fourth share of women diagnosed with cancer in India. Due to the genetic changes in the human body, the regular replacement of old cells with new one could not be done. As a result, the cells grow into a mass which is often called as tumor. But all tumors are not cancerous. Non-cancerous tumors are called Benign as they do not cause much harm and don not spread while their counterpart are referred as Malignant. A malignant tumor can become sizeable at its primary site and affects the normal functioning of the organ .Though there are many carcinogens, we cannot attribute the cancer to a single cause.

Based on the symptoms, the doctors may ask to have screening test. Occasionally a cancer can be discovered by chance. There are many methods used for breast cancer screening like Mammography, MRI and breast ultrasound, Breast needle biopsy, Cyst or fine needle aspiration. Though digital mammographs are reliable, they have con of emitting harmful X-rays. In addition, they are painful for the women undergoing mammography, particularly young women and women with dense breasts. MRIs are expensive. Due to speckle noise, the quality of imaging will gets effected and hence well-trained radiologists are requited for interpretation. Computer Aided Diagnosis (CAD) could be advantageous to assist radiologists in the US-based detection of breast cancer, as they reduce the impact of the operator-dependent nature of US imaging.

Overall, the state-of-the-art methods are not robust, particularly the image processing based approaches, relying on rule based approaches and specific assumptions. As deep learning approaches have proved their accuracy in object detection, it implies that they can also be applicable for lesion detection in breast ultrasound.[1] Deep learning in medical imaging is mostly represented by convolutional networks.

Ultrasound is harmless and easy, and it lets the radiologist to view the images of the various internal organs of the body using sound waves. Ultrasound imaging, is also referred as ultrasound scanning or sonography. It mainly encompasses the use of a probe which acts as a transducer and ultrasound gel placed directly on the skin. The probe emits High-frequency sound waves and are transmitted into the body through the gel. The reflected sound waves are received by the transducer and are processed in a computer to obtain the image. Ultrasound examinations do not use ionizing radiation (as used in x-rays), thus there is no radiation exposure to the patient. Because ultrasound images are captured in real-time, they can show the structure and movement of the body's internal organs, as well as blood flowing through blood vessels.

Ultrasound scan is a noninvasive medical test that helps doctors diagnose and treat medical conditions. Breast Ultrasound imaging outputs a picture of the internal structures of the breast. While examining with ultrasound, the radiologists might use Doppler techniques to assess the whether the blood flow is present or not in any breast mass. Sometimes, such kind of examination provides further information regarding the cause of the mass. Breast Ultrasound imaging is helpful to decide whether the anomaly present is cancerous or not. Ultrasound screening presents some of the abnormalities which are missed during mammography there by it avoids unwanted biopsy. It is observed that majority of the abnormalities investigated with screening breast ultrasound are not cancer (false positives). Ultrasound can be offered as a screening tool for women who are at high risk for breast cancer and unable to undergo an MRI examination and for the expectant mothers or should not be exposed to x-rays .

So, we can determine breast ultrasound scan is

Revised Manuscript Received on April 06, 2019.

S Sri Durga Kameswari, Department of ECE, Sathyabama Institute of Science and Technology, Chennai, India.

S.Sri Durga Kameswari, Department of ECE, GMR Institute of Technology, Rajam, India

Dr. V Vijayakumar, Department of ECE, Sathyabama Institute of Science and Technology, Chennai, India

non-invasive and economically feasible method of screen. Also, it is free from harmful radiations. Biopsy and Fine Needle Aspiration (FNA) are invasive methods which extract tissues and cells to decide about the nature of the tumor. But these are painful methods. Even though the tumor is benign, the patient has to suffer as the doctor cannot decide without biopsy. The role of Bio-medical engineering is to provide dependable and economic diagnosis techniques with less trauma.

In this work, we used breast ultrasound images as the inputs for classification. The main objective is to design deep learning technique which can tell whether the test image is benign or malignant.

For analyzing the breast ultrasound images, image processing was extensively used in the last decade. But, due to the poor image quality of the acquired image, pre-processing techniques need to be applied in order to remove the noise. After that segmentation follows in which the Region of interest will be obtained. This stage is followed by feature selection and classification. In feature selection stage, best features will be extracted among various features and finally in the classification stage, the tumor will be classified as benign or malignant. There are many approaches developed by various researchers in this domain. Some of the popular classification techniques like SVM, Bagging etc.

Neural networks are inspired from biological neurons in the brain which are capable of memory and decision making. Artificial neural networks (ANNs) are interconnection of neurons or nodes. They are the computer or machine style of brain. The impact of ANNs is felt in all domains. In the recent past there is a vast increase in the usage of neural networks. ANNs, as described earlier, are replica of biological brain and we have to teach the network with some data. This process is called training. After training, testing of the neural networks is done to analyze their performance.

Neural networks have several layers of interconnected nodes or neurons. There are many architectures in neural networks based on how the layers are connected, the number of nodes in each layer, the weights and bias used, number of hidden layers and the type of layers. Neural networks transfers the data in the form of input values through the network connections.

Deep neural networks employ many layers in their architecture. These are used lately in many applications due to their adaptability and high accuracy. These networks once trained with a dataset, can produce better outputs. Convolutional Neural Networks (CNNs) are the most prominent architecture of deep learning. CNNs are preferred over ANNs as the latter need more number of input parameters. For example, let us consider an image of size 24x24(576 pixels) and ANN's layer has 100 neurons and 12 output parameters. Now the number of connections in this ANN will be equal to (576 x 100 + 100 bias + 100x12 output neurons + 12 bias)= 58912.The number become even larger for large images. This is the main drawback of ANNs. It is not necessary to have the entire pixels of the input image to represent it. But we can make the image recognizable with small set of pixels too. In CNN, we use a convolutional matrix which is often called as kernel or filter which is applied to the input image. Now again consider the same 24x24 image applied as input to a CNN. Here there will be a decline in the number of connections. If the Kernel size is 11, then the convoluted output matrix size will be equal to 12(24-11+1). So in this case, the number of parameters that are to be trained are only 121(11x11).This kernel is applied or moved throughout the input image. In other words, we can say that the network sees the image with this filter. We can use predefined weights or filters such as horizontal filter, Sobel filter, Schorr filter. However, the CNN can automatically estimate the weights of the filter.

.

II. METHODOLOGY

Based on the symptoms, the physician may endorse for screening. In this work, we considered breast ultra sound images obtained from Mendeley Dataset[2]. Ultra sound images that are used here are pre classified and only the ROI is presented. As the dataset is already classified as Benign and Malignant we can use it for training our CNN. Fig 1 explains work flow.

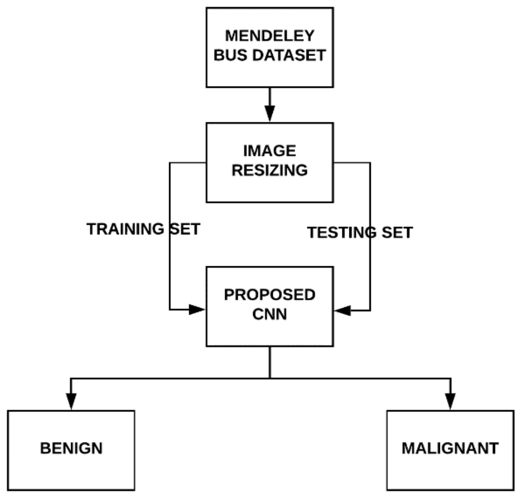

Fig. 1 Work flow of proposed model

A. Training the CNN

Training a NN means providing it the needful data so that it can learn. This dataset has 250 breast ultrasound images in which 150 were malignant and 100 were benign. Here, we have developed a new CNN which is trained from scratch with the available dataset. 90% of the dataset was used as training set. Even though, the acquired dataset is not too large, and we want to continue with it, as it is difficult to find breast ultrasound datasets which are open to all. The images in this dataset are not of uniform size. Hence it is necessary to make all the images of same size otherwise it is hard to train the network. We have resized the data into 72x72 and converted to numpy array. One hot encoding is applied to the labels of the data. In the initial layer of the network, a convolution filter of size 3x3 is applied. 32 filters were used and the Exponential Linear unit as the activation function. Maxpooling (2, 2) is applied with stride of 2. Padding is employed to make the output size of the layer equal to the input size. A dropout rate of 0.3 is used for this layer to minimize the over fitting. In the subsequent two layers, same parameters were used for CNN, but Relu activation function was used. This was followed by an output layer which is the final stage. In this stage, flattening is performed to obtain a single vector and dense layer with 512 connected neurons. The out parameters are only two i.e. benign and malignant. Hence, the output layer has two neurons activated with the softmax function. This model is trained for 50 epochs.

The fig.2 describes the architecture of the proposed CNN

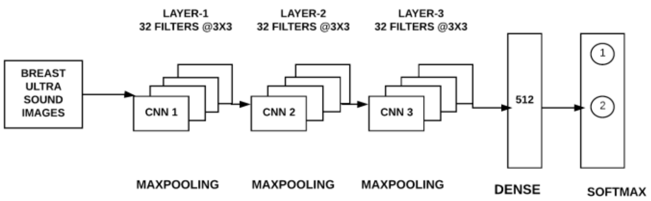

Fig. 2 Architecture of proposed C

B. Testing

Testing is the actual stage during which we can assess the performance of our work. In this work, we have used 10 % of the image dataset for validation. The soft max layer denotes two labels based on the classification done by the CNN, which are *Benign* and *Malignant.*

III. RESULTS AND DISCUSSION

The training and testing is performed on Mendeley Dataset of Breast Ultra sound images that are pre classified as Benign and Malignant. The validation generated an accuracy of 98%. As the data set is not too large there is a chance of over fitting. The normalized confusion matrix is shown in the figure.

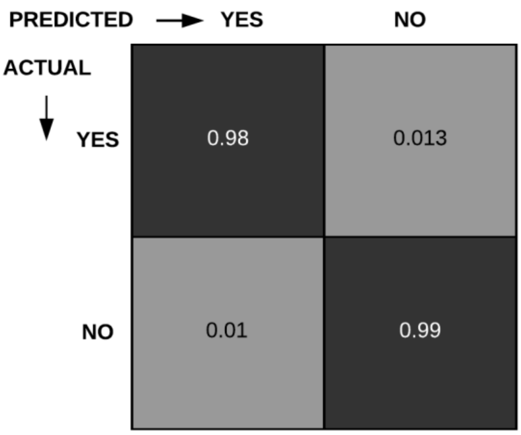

Fig. 3 Confusion matrix of proposed model

Performance

The following parameters indicate the performance of a classifier.

a. **Sensitivity** is indicated by the true Positive Rate. It means the samples indicated as positive or malignant are truly malignant.

b. **Specificity** is indicated by the true negative rate. It means the samples identified as benign are truly benign.

c. **Accuracy** indicates the number of correctly identified samples from the dataset.

The above mentioned parametes are calculated and given in Table1.

Sensitivity	TPR = TP / (TP + FN)	**0.9933**
Specificity	SPC = TN / (FP + TN)	**0.9802**
Accuracy	ACC = (TP + TN) / (P + N)	**0.9880**

Table 1. Performance of the proposed method

IV.CONCLUSION

In this work, we have acquired Breast ultra sound images as our dataset and designed a neural network using CNN layers and performed the classification of the tumors as benign and malignant classes. The results obtained are convincing and this method is reliable. However, the dataset is small which may result in overfitting. In future, we would like to apply the algorithm for a relatively larger dataset and observe the performance of the algorithm.

REFERENCES

1. M. H. Yap *et al.*, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218-1226, July 2018.
2. Rodrigues, Paulo Sergio (2017), "Breast Ultrasound Image", Mendeley Data, v1http://dx.doi.org/10.17632/wmy84gzngw.1
3. H.D.Chenga,JuanShanaWenJua,YanhuiGuoa,LingZhangb "Automated breast cancer detection and classification using ultrasound images: A survey", journal of Pattern Recognition Volume 43, Issue 1, January 2010, Pages 299-317
4. Wei Liyang et al., "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications", Medical Imaging IEEE Transactions on, vol. 24, no. 3, pp. 371-380, 2005.
5. Prabhakar T., Poonguzhali S. (2018) Assessment of Texture Feature Extraction to Classify the Benign and Malignant Lesions from Breast Ultrasound Images. In: Dash S., Naidu P., Bayindir R., Das S. (eds) Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol 668. Springer, Singapore.
6. H. Cheng, J. Shan, W. Ju, Y. Guo, L. Zhang, Automated breast cancer detection and classification using ultrasound images: a survey. Pattern Recogn. **43**, 299–317 (2010).
7. J. Ker, L. Wang, J. Rao and T. Lim, "Deep Learning Applications in Medical Image Analysis," in *IEEE Access*, vol. 6, pp. 9375-9389, 2018.
8. Cho, J., Lee, K., Shin, E., Choy, G., Do, S.: How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? In: ICLR (2016).
9. Benjamin Q. Huynh, Hui Li, Maryellen L. Giger," Digital mammographic tumor classification using transfer learning from deep convolutional neural networks, J. of Medical Imaging, 3(3), 034501 (2016). https://doi.org/10.1117/1.JMI.3.3.034501.
10. Screening Breast Ultrasound: Past, Present, and Future Rachel F. Brem, Megan J. Lenihan, Jennifer Lieberman, and Jessica Torrente American Journal of Roentgenology 2015 204:2, 234-240

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

AUTHORS PROFILE

S Sri Durga Kameswari is currently working as Assistant professor in GMR Institute of Technology, Rajam, AP. She is a Research scholar at Sathyabama Institute of Science and Technology. Her research interests include, signal processing, image processing, biomedical signal processing, deep learning.

Dr. V. Vijayakumar is working as Associate Professor, ECE Department, at Sathyabama Institute of Science and Technology, Chennai. His research interests include High Speed Devices, VLSI Design and signal processing.